GENETIC ORGANIZATION OF CELLOBIOSE TRANSPORT SYSTEM IN *BACILLUS STEAROTHERMOPHILUS*

By

XIAOKUANG LAI

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1994

ACKNOWLEDGEMENTS

I would like to thank Dr. Lonnie O'Neal Ingram, my major advisor, for his guidance, support, motivation, and patience, for my completion of this dissertation. I am extremely grateful to the members of my supervisory committee, Dr. James A. Lindsay, Dr. Francis C. Davis, Dr. John E. Gander, and Dr. Keelnatham T. Shanmugam, for their contributions to my research and to the preparation and review of this manuscript. I must also thank David Beall, Joy Doran, and Lorraine Yomano for their friendship and help during the course of my research.

I wish also to express my deep gratitude to Dr. Richard Hartmann and Mrs. Remy Hartmann for their love and care. Thanks are due to my family for their encouragement and support. Finally, I would like to thank my loving wife, Dongping, and beautiful daughter, Diana, for making my life special.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS ii

LIST OF TABLES v

LIST OF FIGURES vi

ABSTRACT viii

CHAPTERS

1 INTRODUCTION 1

2 REVIEW OF LITERATURE 5
  - Thermophilic Microorganisms 5
  - Thermophilic *Bacillus* 15
  - Bacterial Cellulase System 18
  - Sugar Transport Systems in Bacteria 19
  - Phosphoenolpyruvate-Dependent Phosphotransferase System 20

3 ISOLATION OF THERMOPHILIC CELLULOLYTIC *BACILLUS* STRAINS 41
  - Materials and Methods 42
  - Results 44
  - Discussion 52

4 CLONING AND SEQUENCING OF CELLOBIOSE-SPECIFIC PERMEASE OF PHOSPHOENOLPYRUVATE-DEPENDENT PHOSPHOTRANSFERASE SYSTEM FROM *BACILLUS STEAROTHERMOPHILUS* 55
  - Materials and Methods 56
  - Results 60
  - Discussion 89

5 CLONING AND SEQUENCING OF THE PHOSPHOENOLPYRUVATE-DEPENDENT PHOSPHOTRANSFERASE GENES (*PTSHIT*) FROM *BACILLUS STEAROTHERMOPHILUS* AND HYPEREXPRESSION IN *ESCHERICHIA COLI* 93
  - Materials and Methods 94
  - Results 97
  - Discussion 124

CHAPTERS Page

6 SUMMARY AND CONCLUSIONS 128
LITERATURE CITED 135

BIOGRAPHICAL SKETCH 149

LIST OF TABLES

Table Page

1 Differential properties of thermophilc *Bacillus* . 16

2 Sequenced EI and HPr genes of PTS 24

3 Differential properties of 55°C and 65°C isolates 45

4 Comparison of amino acid sequences in carboxyl terminal portion of alanine racemases from XL-65-6 and other organisms 47

5 Enzymatic activities expressed by isolates capable of growing at 65°C or higher 48

6 Enzymatic activities expressed by isolates capable of growing at 55°C or lower 50

7 Strains and plasmids used in this study 57

8 Comparison of predicted amino acid sequences for *celA, B,* and *D* and *cel'R* from *B. stearothermophilus* to homologous PTS polypeptides from other organisms 70

9 Characteristics of the *cel* operon from *B. stearothermophilus* XL-65-6 74

10 Codon usage of *cel* operon 78

11 *E. coli* strains and plasmids used in this study . 95

12 Comparison of predicted amino acid sequences of *ptsH* and *ptsI* from *B. stearothermophilus* to those from other bacteria 111

13 Characteristics of *ptsHIT* opern 119

LIST OF FIGURES

Figure		Page
1	Organization of *ptsH* and *ptsI* in representative organisms	28
2	Schematic depicition of representative PTS permeases and their energy-coupling proteins found in enteric bacteria, *B. subtilis,* and *R. capsulatus*	32
3	Restriction map of the *B. stearothermophilus* DNA fragment contained in pLOI902	61
4	Nucleotide sequence of DNA fragment from *B. stearothermophilus* contained in pLOI902	64
5	Comparison of deduced amino acid sequences of *cel* operon genes from *B. stearothermophilus* and *E. coli*	72
6	Southern hybridization analysis of *cel* operon . .	81
7	Hydropathy plots of PTS enzymes from different organisms	83
8	Comparision of helical wheel plots of terminal regions from PTS enzymes	86
9	Diagram illustrating the features of recombinant plasmids containing *B. stearothermophilus* DNA . .	99
10	Southern hybridization analysis of *ptsHIT* operon .	102
11	Nucleotide sequence of *B. stearothermophilus* DNA fragment in pLOI1801	104
12	Comparision of deduced amino acid sequences of *ptsH* from *B. stearothermophilus* and other Gram positive bacteria	112
13	Comparison of nucleotide sequence of *ptsH* from *B. stearothermophilus* and from other Gram positive bacteria	116

Figure Page

14 SDS-PAGE of protein extracts from JLT2 recombinants 120

15 Model for cellobiose transport and metabolism in *B. stearothermophilus* 132

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

GENETIC ORGANIZATION OF CELLOBIOSE TRANSPORT SYSTEM IN *BACILLUS STEAROTHERMOPHILUS*

By

Xiaokuang Lai

August 1994

Chairperson: Lonnie O'Neal Ingram
Major Department: Microbiology and Cell Science

A large collection of Gram positive, spore-forming, thermophilic, cellulolytic bacilli were isolated from diverse natural sources. These organisms were characterized and fell into two groups. One group which grew at 55°C or lower was identified as *Bacillus coagulans*. The other group which grew at over 65°C was identified as *Bacillus stearothermophilus*. Both groups of organisms represent a rich source of bacterial genes encoding enzymes for the depolymerization of cellulose to cellobiose and glucose and for the depolymerization of other polysaccharides.

Little was previously known about the mechanisms of cellobiose transport and metabolism in Gram positive bacteria. This was extensively investigated in one of the *B. stearothermophilus* isolates, strain XL-65-6. The cellobiose transport system in this strain consisted of a multicomponents phosphoenolpyruvate-dependent phosphotransferase system (PTS). The three genes encoding the cellobiose-specific permease and the gene encoding a proposed phospho-ß-glucosidase were cloned and sequenced. These were found to be organized as an operon which is present as a single copy on the *B. stearothermophilus* chromosome. The *celB* product forms the major membrane-spanning channel ($EIIC^{Cel}$) for this transport system. Products of *celA* and *celD* are cytoplasmic components of enzyme II ($EIIB^{Cel}$ and $EIIA^{Cel}$). The *celC* gene encodes the glycohydrolase which serves as a cleavage enzyme. It is likely that this enzyme is phospho-ß-D-glucopyranosidase. These genes were functionally expressed in recombinant *Escherichia coli*.

Genes (*ptsH* and *ptsI*) encoding the PTS general proteins, HPr and enzyme I, were also cloned and sequenced from strain XL-65-6. These two genes together with a third, *ptsT*, were also organized as an operon. This *ptsHIT* operon exists as a single copy in the *B. stearothermophilus* chromosome. Deduced amino acid sequences of *B. stearothermophilus ptsH* and *ptsI* are most similar to those from *B. subtilis* although they exhibit a high degree of similarity with homologues from many other bacteria. No gene similar to *ptsT* was found in the current GenBank data base. The exact function of *ptsT* remains undetermined.

Results from these studies provided the basis for a model of cellobiose transport and metabolism in *B. stearothermophilus*.

CHAPTER 1
INTRODUCTION

Thermophiles are a fascinating group of microorganisms which have received considerable interest in recent years because of their potential for biotechnological applications. Research on thermophiles is expanding into many different fields including biochemistry, enzymology, genetics, and molecular biology. The biotechnological interest will further stimulate basic research on this group of microorganism and increase our understanding of thermophiles in the future (Kristjansson and Stetter, 1992).

The genus *Bacillus* is one of the earliest bacteria to be described. Species within this genus have been important as models for studying biological problems, including differentiation, secretion, DNA replication, DNA repair, chemotaxis, genetic transformation, and translation apparatus (Skepecky and Hemphill, 1992). They have also been utilized in a wide range of industrial processes due to metabolic diversity and low reported incidence of pathogenicity (Harwood, 1989).

Interest in cellulose decomposition and in the microorganisms that produce enzyme systems for cellulose depolymerization have been stimulated by the perceived need to exploit cellulose as a renewable source of energy and chemical feedstocks, and to improve the efficiency of digestion of fodder by ruminants (Coughlan and Mayer, 1992). Cellulose is the most abundant carbohydrate in nature. It has been estimated that 7 x 10^{11} metric tons of cellulose exist in primary sources at any one time, and that approximately 4 x 10^{10} tons are replenished each year by photosynthesis (Coughlan, 1985). Unlike fossil fuels, cellulose is clearly a vast and renewable resource. In addition to primary sources, cellulose also represents approximately 50% of the dry weight of many secondary sources of biomass such as agricultural, forest, industrial, and domestic waste. Huge amounts of these materials are also generated each year (Coughlan and Mayer, 1992).

Cellobiose is one of the primary soluble extracellular products of microbial cellulose degradation. Despite the abundance of cellobiose in the environment, little is known about its entry into cells and its intracellular metabolism. Many Gram positive and negative bacteria have been reported to contain cell-associated ß-glucosidases (Coughlan and Mayer, 1992). *Cellulomonas* species have cellobiose phosphorylases which are presumed to be involved in intracellular cellobiose metabolism (Coughlan and Mayer, 1992). In *E. coli*, two cryptic operons have been described at a molecular level which provide the potential to transport cellobiose into the cell by a phosphoenolpyruvate-dependent phosphotransferase system (PTS) and to cleave cellobiose into monomer (Hall and Xu, 1992; Parker and Hall, 1990). Both operons have different gene organizations but potentially encode a cellobiose-specific enzyme II complex and a phospho-ß-glucosidase. This is the only organism for which cellobiose uptake has been defined. But this organism typically does not have a functional cellobiose transport system.

Different organisms often transport the same sugars by alternative mechanisms. For example, *E. coli* and *Bacillus* species use non-PTS to take up lactose into the cell while *Staphylococcus* and *Streptococcus* species initiate the metabolism of this sugar by the PTS (Saier and Yamada, 1987). However, no information is available about the uptake of cellobiose in any Gram positive bacteria.

PTS is a complex system which is used to transport a variety of sugars in different bacteria. This system consists of several interacting proteins, including cytoplasmic and membrane proteins (Postma et al., 1993). In addition to its primary function of carbohydrate translocation, the PTS also has a variety of other important physiological functions in bacterial cells. One such function is to regulate the order of sugar utilization when mixtures of sugars are present. For example, in media containing sugars transported by the PTS (PTS-sugar) and sugars transported by other system (non-PTS sugar), PTS-sugars are utilized prior to induction of the catabolic systems for remaining sugars (diauxic growth). The PTS can also function as a signal transduction system through which bacteria respond to the presence of PTS-sugars by positive chemotaxis. The PTS may also serve as a link between carbon and nitrogen metabolism in Gram positive bacteria. One domain of the *B. subtilis levR* gene product, a protein involved in the regulation of PTS levanase operon, is homologous to a part of NifA and NtrC proteins, which are activators of σ^{54}-dependent transcription (Débarbouillé et al., 1991). σ^{54} is required for the transcription of a number of genes important for glutamate production, especially under the conditions of nitrogen limitation.

The primary aim of this dissertation is to investigate the genetic organization of cellobiose uptake system in a thermophilic species of *Bacillus*. The specific objectives were as follows:

1. To isolate and identify thermophilic cellulolytic *Bacillus* strains from nature.
2. To characterize cellobiose uptake genes at a molecular level from one isolate.
3. To construct cellobiose-utilizing strains of *E. coli* by introducing the *Bacillus* genes into *E. coli*.

CHAPTER 2
REVIEW OF LITERATURE

Thermophilic Microorganisms

Definition of Thermophilic Microorganisms

Temperature is one of the most important variables in our environment. The classification of living organisms based on their relation to temperature has therefore always been considered as one of the most basic elements of biological systematics. Microorganisms are frequently divided into three broad classes on the basis of temperature ranges for growth, that is psychrophiles, mesophiles, and thermophiles (Neidhardt et al., 1990). Brock (1986) has suggested a definition of a thermophile boundary at 55°C to 60°C based on two main arguments: first, temperatures lower than 50°C are widespread on earth, whereas temperatures greater than 55°C to 60°C are much rarer in nature. Second, 60°C is the upper temperature limit for eucaryotic life. However, bacteria that grow at 50°C or above are traditionally called thermophiles (Kristjansson and Stetter, 1992; Neidhardt et al., 1990).

The upper temperatures for different thermophilic bacteria vary considerably. In general, nonphotosynthetic organisms are able to grow at higher temperatures than photosynthetic forms; structurally less complex organisms can grow at higher temperatures than more complex organisms. The upper growth temperature for cyanobacteria and other phototrophic bacteria is 70°C to 73°C, for organotrophic bacteria is over 90°C, while archaebacteria can grow at over 100°C. However, not all organisms from these groups are thermophiles. Usually only a few species or genera are able to function successfully near the upper temperature limits (Brock and Madigan, 1991).

General Physiological Characteristics of Thermophiles

Thermophilic bacteria exhibit a wide range of capabilities for energy and nutrition metabolism: phototrophy and chemotrophy, autotrophy and heterotrophy, aerobiosis and anaerobiosis. Examples of thermophiles capable of phototrophic growth include 16 cyanobacteria species and some purple and green bacteria (Brock and Madigan, 1991). Although these organisms use light as an energy source, most thermophiles are chemotrophs, using either inorganic or organic chemicals as energy sources. Thermophiles that grow autotrophically include many of the methanogens, *Bacillus schlegelii* and *Clostridium thermoautotrophicum*. However, like the majority of thermophilic bacteria, these organisms also grow heterotrophically (Cato et al., 1986; Claus and Berkeley, 1986). Many thermophiles, including most of the thermophilic members of the genera *Bacillus*, *Clostridium*, *Thermomicrobium*, and *Thermus* only grow heterotrophically. Heterotrophic thermophiles utilize a wide range of carbon sources for growth, including carbohydrates, protein, lipids, organic acids, alcohols, aromatic compounds as well as many other compounds (Sundaram, 1986). Thermophilic *Bacillus* strains have been isolated which are able to degrade methanol (Dijkhuizen et al., 1988; Al-Awadhi et al., 1989), ethanol (Al-Awadhi et al., 1989), phenol, and cresol (Buswell and Twomey, 1975; Buswell, 1975). Examples of aerobic thermophiles include the genus *Thermus* (Williams, 1992), and thermophilic species in the genus *Bacillus* (Claus and Berkeley, 1986). For these organisms, oxygen is involved in the energy-generating step, serving as a terminal electron acceptor. In contrast to this group, some thermophiles, such as thermophilic species in the genus *Clostridium*, are anaerobes. These organisms produce energy through anaerobic fermentative pathways.

Adaptation of Thermophiles to High Temperature

Considerable information has been obtained about the ecology, physiology, metabolic capacities, and biochemical and physical properties of cellular components of thermophiles. It is clear that the ability to grow at high temperature cannot be explained by a single mechanism but, rather, that thermophiles appear to use a variety of mechanisms (Welker, 1993).

The content of guanine and cytosine (G+C) often correlates with maximum growth temperature, although *Clostridium* species are an exception (Welker, 1976). The DNA of thermophiles showed a consistently higher G+C content than that of DNA from mesophiles. Thermal melting profiles demonstrated that DNA from thermophiles had higher melting temperatures than DNA from mesophiles. The greater stability to thermally-induced strand separation is attributable to the more extensive hydrogen bonding that occurs with a higher G+C content.

Stability of the cellular membrane is a determinant of upper temperature limit for growth of organisms (Sundaram, 1986). The structure of cell membrane of thermophilic bacteria, like those of other biological membranes, conforms to the model of a lipid bilayer. However, the lipids in thermophiles tend to have a greater abundance of fatty acids with higher melting points than do the lipids of mesophiles. A significant property of the phospholipid bilayer of the bacterial membrane is that it can undergo thermotropic, reversible transition between an ordered, rigid gel or solid phase and a more fluid liquid-crystalline phase. The change from gel to liquid-crystalline phase involves the cooperative melting of the hydrocarbon chains in the interior of the bilayer, and this phase transition behavior may set the minimal and maximal temperatures for growth (Sundaram, 1986). The stability of proteins also determine the maximal growth temperature (Sundaram, 1986). A large number of proteins from thermophiles have been isolated in pure form. Generally, these proteins are more thermostable than their homologous counterparts from mesophiles (Sundaram, 1986). Amelunxen and Lins (1968) compared the thermostability of 11 enzymes from the thermophile *Bacillus stearothermophilus* and the mesophile *Bacillus cereus*. They found that nine of these enzymes from *B. stearothermophilus* were more thermostable than the corresponding enzymes from *B. cereus*. However, pyruvate kinase and glutamate-oxaloacetate transaminase from both bacteria exhibited similar rates of inactivation at 70°C.

Thermophilic adaptation has been correlated with preferential alterations of the amino acid composition in terms of 'traffic rules' governing gross amino acid compositions of mesophilic and thermophilic proteins. Mozhaev and Martinek (1984) concluded that many thermophilic proteins are deficient in polar amino acids, mostly serine and threonine. Substitution of the inner serine and threonine for nonpolar amino acid residues was proposed as a mechanism to stabilize proteins to high temperature. DiRuggiero et al. (1993) reported that thermostable glutamate dehydrogenase (GDH) from hyperthermophilic Archaeon ES4 (optimal growth temperature 98°C and maximum growth temperature 110°C) had a relatively high hydrophobicity and a low number of sulfur-containing residues compared with mesophilic GDHs. However, after statistically comparing dihydrofolate reductases from 22 different organisms, Böhm and Jaenicke (1994) concluded that simple 'traffic rules' were insufficient to explain the mechanisms of thermophilic adaptation or to predict the stabilization or destabilization of proteins.

Indeed proteins from thermophiles are rather similar to their mesophilic counterparts in most respects, including size, subunit structure, gross higher order structural parameters, and modulation of activity by metal ions and other effectors (Amelunxen and Murdock, 1978). Attempts to transfer heat stability and heat lability by mixing cell-free extracts from thermophiles and mesophiles failed to change the original properties of individual enzymes (Amelunxen and Lins, 1968; Koffler and Gale, 1957). These results established that the functioning of proteins at high growth temperatures of thermophiles is not merely dependent on the presence of stabilizing factors in the cellular environment of the thermophile and that intrinsic structural stability of proteins is the more important basis for thermophily (Sundaram, 1986).

Amelunxen and Murdock (1978) have suggested that the structural stabilization of thermophilic proteins is primarily achieved through extra hydrogen bonds, ionic interactions and apolar interactions in certain critical parts of the protein molecules. After comparing the structures of triosephosphate isomerase from thermophilic bacteria, *Moraxella sp.*, and mesophilic chicken, Rentier-Delrue et al. (1993) concluded that the adaptability of this enzyme to high temperature appeared to be favored by better stabilizing residues for the helix dipole as well as better helix-forming residues.

Also, there is evidence that thermostability of a protein can be altered by a single amino acid change. The thermostability of *E. coli* ribonuclease HI was increased by replacement of single amino acid residues (Ishikawa et al. 1993). Crystal structures of these mutant proteins revealed that only changes in the local structure around each mutation site are essential for the increase in thermostability. After His65 was replaced by Pro, the mutant protein was stabilized because of a decrease in the entropy of the unfolded state, without a change in the native backbone structure. After Lys95 was replaced by Gly, the left-handed backbone structure in the typical 3:5 type loop was eliminated. The mutant protein with a replacement of Lys95 by Asn was stabilized due to formation of a hydrogen bond between the side-chain Nδ-atom of the asparagine residue and the main-chain carboxyl oxygen within the same residue.

Inherent structural stability alone may not be sufficient to ensure the survival and functioning of proteins in thermophilic cells. Thermophilic proteins display a rather broad spectrum of stability *in vitro*, suggesting that they are not all equally stable. Many enzymes from thermophiles are stabilized *in vitro* by supplements, such as substrates, metal ions, allosteric effectors, and electrolytes, all of which are present in the growing bacterial cells (Sundaram, 1986).

Glutamine synthetase from *Bacillus caldolytics* was demonstrated to be stabilized by the binding of metal ions and substrate (Merkler et al., 1988). *In vitro* stability of glyceraldehyde-3-phosphate dehydrogenase from *Bacillus coagulans* was enhanced by increasing the ionic strength to 1.8 M with many neutral salts (McLinden et al., 1986). Calcium ions confer stability to several extracellular enzymes, such as α-amylase (Yutani, 1976), thermolysin (Tajima et al., 1976), and proteases (Sidler and Zuber, 1977).

Insights into the molecular mechanisms of thermophily have been hampered because little is known about the genetic and molecular basis of thermophily. This gap is directly related to lack of a genetic exchange systems for the manipulation and genetic analysis of most thermophilic species (Welker, 1993). Lindsay and Creaser (1975) reported that after a mesophilic *B. subtilis* strain was transformed using DNA isolated from a obligately thermophilic *B. caldolyticus* strain, the resulting recombinant of *B. subtilis* strain was able to grow above 70°C. These investigators suggested that thermophily was controlled by a small number of genes acting at the ribosomal or tRNA level affecting translation resulting in expression of more thermostable enzyme variants. An α-glucosidase purified from the high-temperature-growth transformant had been confirmed to have changes in the amino acid composition. Although this enzyme had much higher optimum growth temperature than its counterpart in its mesophilic parent, many physiological characteristics remained unchanged (Krohn and Lindsay, 1991). Similar results have been reported by Droffner and Yamamoto (1985). Most recently, Ståhl (1991) isolated 3 plasmids from thermophilic *B. stearothermophilus*, which could be involved in thermophily since strains cured of these plasmids were no longer obligate thermophiles. Furthermore, when these plasmids were transformed into a mesophilic *B. subtilis* strain, the resulting transformants shifted their optimum growth temperature about 10°C higher, and about 6°C higher for the maximum growth temperature. Based on these experiments, Ståhl concluded that Lindsay and Creaser's results (1975) might be due to the presence in the donor strain of plasmids involved in thermophily.

Evolution of Thermophilic Microorganisms

Thermophilic bacteria are distributed among all main groups in bacterial genetic tree. However, it is interesting to note that they are usually the oldest types in their respective groups (for example *Thermotoga*, *Thermomicrobium*, *Chloroflexus*, and *Thermous*), and even the oldest of the bacterial kingdom. These results suggest that thermophily arose very early in the evolution of bacteria. Indeed, the ancestral bacterium might have been a thermophile (Kristjansson and Stetter, 1992).

Applications of Thermophilic Microorganisms

Thermophiles offer some major advantages for biotechnology. The most attractive attribute of thermophiles, from a biotechnological point of view, is that they produce enzymes capable of catalyzing biochemical reactions at temperatures markedly higher than those of conventional organisms. The main advantages of increased temperature are generally higher reaction rates, higher solubility of most chemicals, and increased fluidity and diffusion rates. High temperature also prevents contamination from mesophilic microorganisms. In addition, many enzymes from thermophiles are more stable at conventional temperatures, thus prolonging the shelf life of commercial products (Brock, 1986). Interest in thermostable enzymes from thermophiles has grown. The variety of thermostable enzymes used in industrial application has been steadily increasing, mainly as replacements for thermolabile enzymes in existing processes. For example, thermostable amylase from *B. licheniformis* and *B. stearothermophilus* have replaced thermolabile amylase from *B. subtilis* (Zamost et al., 1991). The main applications for thermostable enzymes have been starch liquification using amylases and proteases for food processing and detergents (Zamost et al., 1991). Other thermostable enzymes with a wide range of characteristics remain to be exploited and applied. The application of thermophiles is not limited to industrial enzyme processes. There exist many other potential applications, such as fermentation, waste treatment, and microbial leaching (Brock, 1986).

Thermophilic *Bacillus*

In Bergey's Manual of Systematic Bacteriology (Claus and Berkeley, 1986), the thermophilic, aerobic or facultative anaerobic, spore-forming, Gram positive rods are placed in the species *B. acidocaldarius, B. schlegelii, B. stearothermophilus, B. coagulans*. and *B. licheniformis*. The first three species are able to grow at 65°C or higher, and the later two are able to grow at 55°C. The differential properties of these organisms are listed in Table 1.

B. acidocaldarius was first isolated by Darland and Brock (1971) from hot springs in Yellowstone and from fumaroles in Hawaii. This organism appears to be restricted to thermal springs and the surrounding soils (Priest, 1989). It has a temperature range between 45°C and 70°C and a pH range between pH 2 and pH 6 (Darland and Brock, 1971). One of the unique features of *B. acidocaldarius* is its fatty acid profile. Like other *Bacillus* spp., it possesses MK-7 as a major menaquinone component, but in addition it also possesses the unique fatty acid ω-cyclohexane (Minnikin and Goodfellow, 1981).

B. schlegelii is a thermophilic hydrogen-oxidizing organism. It was isolated by Aragno (1978) from the superficial layer of the sediment of a small eutrophic lake in Switzerland and was described by Schenk and Aragno (1979). It is an obligate aerobe. The Gram reaction is variable, but cell wall structure of this organism is typical of a Gram positive bacteria (Aragno, 1992). The optimal temperature for

Table 1. Differential Properties of Thermophilic *Bacillus*

Properties[a]	*Bco*	*Bli*	*Bac*	*Bsc*	*Bst*
Growth at					
30°C	+	+	-	-	-
55°C	+	+	+	+	+
65°C	-	-	+	+	+
Growth in					
5% NaCl	-	+	ND	-	d
7% NaCl	-	+	ND	-	-
Acid from					
D-Glucose	+	+	+	-	+
Hydrolysis of					
Casein	d	+	ND	-	d
Gelatin	-	+	ND	-	+
Starch	+	+	+	-	+
Growth at					
pH 6.8	+	+	-	+	+
Autotrophic with					
$H_2 + CO_2$ or CO	-	-	-	+	-

Source: Claus and Berkeley, 1986

Note: *Bco*: *Bacillus coagulans*
Bli: *Bacillus licheniformis*
Bac: *Bacillus acidocaldarius*
Bsc: *Bacillus schlegelii*
Bst: *Bacillus stearothermophilus*

[a] +: 90% or more strains are positive
-: 90% or more strains are negative
d: 11-89% of strains are positive
ND: no data available

growth is about 70°C, with no growth at 37°C or 80°C. The optimal pH for growth is between pH 6 to pH 7 (Schenk and Aragno, 1979). *B. schlegelii* is facultatively chemolithotrophic using either H_2 as the electron donor and CO_2 as the carbon source or CO to serve both requirements (Claus and Berkeley, 1986).

B. stearothermophilus is a highly heterogeneous species. Wolf and Sharp (1981) subdivided this species into three groups, while Bartolomeo et al. (1991) clustered this species into four groups on the basis of morphological, physiological, and biochemical characteristics as well as the DNA base compositions of 133 thermophilic *Bacillus* strains.

B. coagulans is an important food spoilage agent involved in the coagulation of canned milk and flat souring of carbohydrate containing canned foods through the production of high concentration of L-(+)-lactic acid (Frazier, 1958). It was originally isolated from spoiled canned milk by Hammer in 1915. This organism is considered to be a facultative thermophile, growing well at 45°C to 55°C with some strains growing up to 65°C (Wolf and Sharp, 1981). Early studies indicated considerable heterogeneity among members of this species with variations in morphology of both cells and spores (Smith et al., 1952; Sei et al., 1978). Wolf and Barker (1968) divided this species into two distinct physiological types. On the basis of a more recent study by Bartolomeo et al. (1991), this species was further divided into three groups.

Although some investigators have proposed species names for certain thermophilic *Bacillus* strains, These strains have not been recognized as species in Bergey's Manual (Claus and Berkeley, 1986).

Bacterial Cellulase System

A complete cellulase system is defined as that which can catalyze extensive hydrolysis of crystalline cellulose (Coughlan and Mayer, 1992). Such systems require the concerted action of several different types of enzymes, among which mainly include: (a) endoglucanases (EC 3.2.1.4, CMCase), (b) cellobiohydrolases (EC 3.2.1.91, exoglucanase), and (c) ß-glucosidases (EC 3.2.1.21). The endoglucanase cleaves ß-glucosidic bonds randomly in the internal region of the cellulose molecules. Typically, these enzymes are inactive against crystalline cellulose although they may yield a small amount of reducing sugar by hydrolysis of the amorphous regions of crystalline substrates. This group of enzymes, generally, is active against acid-swollen amorphous cellulose, soluble derivatives of cellulose such as carboxyl methyl cellulose (CMC), and cellooligosaccharides. Cellobiohydrolases attack cellulose molecules stepwise from the nonreducing ends, liberating soluble cellobiose units. These enzymes can act, albeit slowly, against crystalline substrates such as Avicel. They are also active against amorphous celluloses and cellooligosaccharides, but are inactive against cellobiose or substituted soluble celluloses such as CMC. The ß-glucosidases hydrolyze cellobiose to glucose and catalyze the removal of glucose from the low molecular weight cellooligosaccharides, but they are inactive against crystalline or amorphous cellulose (Coughlan and Mayer, 1992; Wood and Bhat, 1988). In addition to the above three major types of cellulases, many other types of enzymes, such as phospho-ß-glucosidase, cellobiose phosphorylase, cellodextrin phosphorylase, may also play roles in various organisms (Coughlan and Mayer, 1992; El Hassouni et al., 1992).

At least 35 bacterial genera have been reported to contain cellulolytic species, many of which can grow on cellulose and produce enzymes that catalyze the degradation of soluble derivatives of cellulose or the amorphous regions of otherwise crystalline cellulose. However, very few bacteria synthesize the complete enzyme system needed for the extensive hydrolysis of the crystalline cellulose in nature (Coughan and Mayer, 1992).

Sugar Transport Systems in Bacteria

Bacteria employ various transport systems for uptake of different sugars. In *E. coli*, almost all of the transport mechanisms (including facilitated diffusion, shock-sensitive active transport, and secondary active transport) participate in the uptake of one or more sugars or sugar alcohols

(Neidhardt et al., 1990). Glycerol enters *E. coli* by facilitated diffusion with the aid of a glycerol-specific transmembrane protein. This mechanism does not concentrate the substrate within the cytoplasm. Maltose enters *E. coli* by a shock-sensitive system. This system is made up of four distinct proteins, a binding protein, two transmembrane proteins, and an energy-transducing protein. When maltose enters the periplasm, it binds tightly to the binding protein; the binding protein then passes maltose to sites on the transmembrane proteins; the energy-transducing protein hydrolyze a high-energy phosphate bond of ATP, the energy of which changes the conformation of transmembrane proteins such that the bound maltose is released on the membrane's inner surface. Lactose enters *E. coli* by a proton symport and melibiose by a sodium symport. Symport is the transport of two substrates simultaneously in the same direction by a single carrier: if one of these substrates flows down its concentration gradient, the other flows with it. Glucose enters *E. coli* by a phosphoenolpyruvate-dependent phosphotransferase system (PTS), which will be described in more detail in next section. Interestingly, the uptake of a particular sugar does not follow a pattern: for example lactose is taken into *E. coli* by proton symport but into *Staphylococcus aureus* by a PTS.

Phosphoenolpyruvate-Dependent Phosphotransferase System

Bacteria take up many sugars into the cell by a complex system, phosphoenolpyruvate-dependent phosphotransferase system (PTS). This system consists of several interacting proteins and has a variety of physiological functions in bacterial cells in addition to sugar transport (Postma et al., 1993).

The PTS was discovered in cell extracts of *E. coli* by Kundig et al. (1964) as a system which catalyzes the phosphorylation of a number of carbohydrates with PEP as the phosphoryl donor. This system consists of two general energy-coupling proteins, enzyme I (EI) and HPr, as well as a sugar-specific permease, commonly referred as the enzyme II complex (EII) (Saier and Reizer, 1992). The general reaction scheme of the system is illustrated as following (Postma et al., 1993; Saier and Reizer, 1992):

$$\text{P-enolpyruvate} + \text{EI} \rightleftharpoons \text{P-EI} + \text{pyruvate} \tag{1}$$

$$\text{P-EI} + \text{HPr} \rightleftharpoons \text{P-HPr} + \text{EI} \tag{2}$$

$$\text{P-HPr} + \text{EIIA(domain or protein)} \rightleftharpoons \text{P-EIIA} + \text{HPr} \tag{3}$$

$$\text{P-EIIA} + \text{EIIB (domain or protein)} \rightleftharpoons \text{P-EIIB} + \text{EIIA} \tag{4}$$

$$\text{P-EIIB} + \text{EIIC (domain or protein)} \rightleftharpoons \text{P-EIIB} + \text{EIIA} \tag{5}$$

$$\text{P-EIIC} + \text{carbohydrate}_{(out)} \rightarrow \text{EIIC} + \text{carbohydrate-P}_{(in)} \tag{6}$$

In most cases, EI and HPr are soluble, cytoplasmic proteins that participate in the phosphorylation of all PTS carbohydrates in a given organism, and thus are referred to as general PTS proteins (Postma et al., 1993).

Enzyme I

EI has been purified from a variety of organisms (Postma et al., 1993). The monomer molecular weight ranges from about 64,000 in *E. coli* and *S. typhimurium* to about 85,000 in *Staphylococcus aureus*. However, evidence suggests that autophosphorylation of EI (reaction I) requires a dimeric form of the protein (Kukuruzinska et al., 1982; Misset et al., 1980; Weigel et al., 1982). The EIs from most organisms consist of identical monomers that self-associate to form dimers. The exception is the EI from *Mycoplasma capricolum*, which is a tetramer composed of three different subunits (Meadow et al., 1990). Phosphorylation occurs at the N-3 position of a histidyl residue of EI during autophosphorylation with PEP (Alpert et al., 1985; Weigel et al., 1982). Studies on *S. typhimurium* EI shows that this enzyme requires Mg^{2+}, Mn^{2+}, or other divalent cations for its autophosphorylation by PEP but not for the transfer of the phosphoryl group to HPr (Weigel et al., 1982).

The kinetics of the reaction of EI have been well studied and the data have extensively been reviewed (Postma et al., 1993; Meadow et al., 1990). According to the studies on EIs from enteric bacteria, EI associates at room temperature to form an active dimer of identical subunits, which is phosphorylated on each subunit in the presence of Mg^{2+} and PEP. The P-EI then dissociates, especially at lower temperatures, to the monomers which then phosphorylate HPr. Reassociation of dephosphorylated subunits is apparently the rate-limiting step for this whole cycle. This enzyme exhibits a ping-pong Bi Bi mechanism and catalyzes the isotope exchange reactions expected for an enzyme that forms a covalent intermediate. The apparent K_m values range from 0.2 to 0.6 mM for PEP and 4 to 30 μM for HPr.

The gene encoding EI (*ptsI*) has been cloned and sequenced from 3 Gram negative organisms, *E. coli* (Saffen et al., 1987), *S. typhimurium* (LiCalsi et al., 1991), and *Alcaligenes eutrophus* (Pries et al., 1991), and from 4 Gram positive organisms, *Bacillus subtilis* (Reizer et al., 1993), *Staphylococcus carnosus* (Kohlbrecher et al., 1992), *Streptococcus mutans* (Boyd et al., 1994), and *Streptococcus salivarius* (Gagnon et al., 1992) (Table 2). In all cases, there is a high degree of sequence similarity around and including the phosphorylated histidine residue (Postma et al., 1993). The deduced amino acid sequences of EIs from *E. coli* and *S. typhimurium* differ in only 16 residues, consistent with the fact that these proteins are functionally interchangeable in these organisms (Postma et al., 1993). However, these EIs and those from Gram positive bacteria often substitute poorly for one another in PEP-dependent phosphorylation of heterologous HPr (Postma et al., 1993). *Rhodobacter capsulatus* produces a multiphosphoryl transfer protein which contains the functional and structural equivalents of EIIA, HPr, and EI in a single polypeptide chain (Wu et al., 1990).

Table 2. Sequenced EI and HPr genes of the PTS

Proteins	No. amino acids	M_r	References
HPr			
Bacillus subtilis	88	9,121	Gonzy-Treboul et al., 1989
Enterococcus faecalis	88	9,449	Deutscher et al., 1986
Staphylococcus carnosus	88	9,511	Eisermann et al., 1991
Staphylococcus aureus	87	9,367	Reizer et al., 1988
Streptococcus mutans	87	8,936	Boyd et al., 1994
Streptococcus salivarius	87	8,901	Gagnon et al., 1993
Alcaligenes eutrophus	89	9,476	Pries et al., 1991
Escherichia coli	85	9,109	Saffen et al., 1987
Klebsiella pneumoniae	85	9,119	Titgemeyer et al, 1990
Mycoplasma capricolum	89	9,418	Zhu et al., 1993
Salmonella typhimurium	85	9,109	Byrne et al., 1988
Psuedo-HPr			
E. coli	376	39,647	Richterich et al., 1994
S. typhimurium	376	39,593	Geerse et al., 1989
Enzyme I			
B. subtilis	550	63,076	Reizer et al., 1993
S. carnosus	574	63,369	Kohlbrecher et al., 1992
S. mutans	577	63,354	Boyd et al., 1994
S. salivarius	577	62,948	Gagnon et al., 1992
A. eutrophus	592	65,207	Pries et al., 1991
E. coli	575	63,489	Saffen et al., 1987
S. typhimurium	575	63,368	LiCalsi et al., 1991
Multiphosphoryl Transfer Protein			
Rhodobacter capsulatus	827	86,393	Wu et al., 1990

In addition, two other enzymes, pyruvate phosphate dikinase (Pocalyko et al., 1990; Wu et al., 1990) and PEP synthase (Niersbach et al., 1992), were found to have regions of sequence which are similar to EI. These included the region containing the phosphorylated histidine residue.

Histidine-Containing Protein (HPr)

HPr, the second general protein involved in the phosphoryl transfer reaction sequence, is a small monomeric protein with a molecular weight of 9,000 to 10,000. Its small size and heat stability have aided its purification from a variety of organisms (Postma et al., 1993). The phosphoryl group from P-EI is transferred to the N-1 position of a histidine residue (His^{15}) in HPr. All HPrs so far studied from Gram positive bacteria can also be phosphorylated by an ATP-dependent HPr kinase at a serine residue (Ser^{46}) as part of an important regulatory modification (Meadow et al., 1990; Reizer et al., 1993). This regulatory phosphorylation also occurs in HPr from *M. capricolum*, an unusual bacterium (Zhu et al., 1993). The region around Ser^{46} is conserved in all sequenced HPrs of Gram positive bacteria. The Ser^{46}-phosphorylated HPr is not able to transfer its phosphate group to a sugar, and the PEP-dependent phosphorylation at His^{15} of a Ser^{46}-phosphorylated HPr is 600-fold slower than the phosphorylation of a non-phosphorylated HPr (Reizer et al., 1993). In *B. subtilis*, this regulatory phosphorylation is involved in catabolite repression (Deutscher et al., 1994; Reizer et al., 1993). Ser^{46} is also present in HPrs from Gram negative bacteria but is not phosphorylated, and the corresponding HPr kinase is absent in these organisms. The region surrounding Ser^{46} in Gram negative organisms shows little sequence identity with those from Gram positive organisms (Meadow et al., 1990; Reizer et al., 1993).

The three-dimensional structures of HPr from both Gram negative and Gram positive bacteria have been investigated by several different methods: two-dimensional nuclear magnetic resonance (NMR) and three-dimensional NMR of HPr in solution and X-ray crystallography (Postma et al., 1993). The structure of the *E. coli* HPr determined by NMR studies consists of four ß-strands that form a single, pleated antiparallel ß-sheet on one face of the protein and three α-helices that lie in a plane parallel to the ß-sheet. According to the X-ray structure of the *E. coli* HPr, however, the four ß-strands exist in two pairs removed from each other and α-helices lie in different planes and orientations, whereas the region containing the phosphorylated His residue is the same in both studies (Postma et al., 1993). The *B. subtilis* HPr has also been investigated by X-ray crystallography and NMR. The overall topology of three dimensional structure of *B. subtilis* HPr derived from NMR and X-ray crystallography is very similar. The secondary-structure folding topology of this protein is the classical topology which has been described as an open-face ß-sandwich.

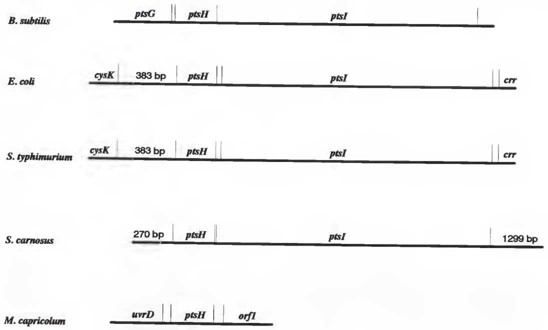
It is formed by four antiparallel ß-strands with three apposed α-helices packed on top of the ß-strands. Both His^{15} and Ser^{46} are located on the surface of the protein with their side chains capping the N-termini of the first and second α-helices, respectively (Chen et al., 1993).

The primary amino acid sequences of HPr from 11 bacteria have been determined (Table 2). Those from *E. coli* and *S. typhimurium* are identical (Byrne et al., 1988), and only one conservative exchange from *K. pneumoniae* (Titgemeyer et al., 1990). All HPrs share a striking similarity around the active site His residue, and all have this active site His residue at position 15. All HPrs also have an Arg residue at position 17 (Meadow et al., 1990). HPrs from Gram positive bacteria, although very similar to each other, do differ considerably in amino acid sequences from the Gram negative bacterial HPrs outside of the region containing the phosphorylated His residue (Postma et al., 1993).

Genetic Organization of *ptsH* and *ptsI*

The arrangements of *ptsH* and *ptsI* on genome are the same in most of bacteria with the exception of *M. capricolum* (Fig 1). In most bacterial species characterized thus far, these two genes are always adjacent to each other and form an operon with *ptsH* on the 5' end. The *M. capricolum ptsH* gene, however, is part of a monocistronic operon that is situated between two ORFs unrelated to the function of PTS (Zhu et al., 1993). The genes on two flanking regions of *ptsHI* operon are

Figure 1. Organization of *ptsH* and *ptsI* in representative organisms


B. subtilis
ptsG
ptsH
ptsI
E. coli
cysK
383 bp
ptsH
ptsI
crr
S. typhimurium
cysK
383 bp
ptsH
ptsI
crr
S. carnosus
270 bp
ptsH
ptsI
1299 bp
M. capricolum
uvrD
ptsH
orf1

vary in different organisms. In some organisms, these regions contain PTS related genes, and in other organisms, genes unrelated to PTS functions are present in these regions. For example, the upstream region of *B. subtilis ptsH* contains the gene encoding glucose permease (*ptsG*) (Zagorec and Postma, 1992), but in *E. coli* and *S. typhimurium*, this region encodes *O*-acetylserine (thiol)-lyase A (*cysK*), an enzyme involved in the synthesis of L-cysteine from *O*-acetyl-L-serine and sulfide (Byrne et al., 1988).

Enzyme II Complexes

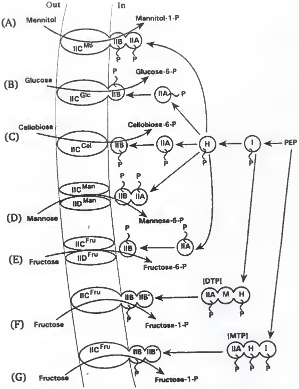
The sugar-specific PTS permeases (EII) are the most diverse PTS proteins (Meadow et al., 1990). They may consist of one, two, three, or four distinct polypeptide chains (Saier and Reizer, 1992). EII proteins found within a single species often show little sequence similarity (Meadow et al., 1990). However, regardless of the number of polypeptide chains and the orders of these chains, most EIIs share several conserved features. First, they generally have similar molecular weights of about 68,000, corresponding to about 635 total amino acyl residues (Saier et al., 1988). Second, there are always at least three well-recognized functional domains: a hydrophobic transmembrane domain which binds and transports the sugar substrate (EIIC), a hydrophilic enzyme III-like domain which possesses the first phosphorylation site (always a histidyl residue) (EIIA), and a second hydrophilic protein or protein domain which possesses the second phosphorylation site (either a cystidyl residue or a histidyl residue) (EIIB) (Fig. 2) (Saier and Reizer, 1992). The activity of EII may be controlled by HPr Ser^{46} phosphorylation via the ATP-dependent kinase in Gram positive bacteria (Ye et al., 1994b).

Regulation of PTS Protein Synthesis

The synthesis of PTS proteins is highly regulated. In enteric bacteria, expression of the *ptsH, ptsI* and *crr* increases about threefold during growth on PTS sugars, especially glucose, when compared to growth on lactate (Mattoo and Waygood, 1983; Stock et al., 1982). Anaerobic growth conditions also favor higher expression (Lengeler and Steinberger, 1978). Transcription of the *ptsHI* operon is also enhanced by unphosphorylated EIICB^{Glc} (De Reuse and Danchin, 1988; De Reuse and Danchin, 1991). These three genes (*ptsH, ptsI, crr*) constitute an operon, but three major mRNA species have been found, composed of those including entire three genes, *ptsH* only, and *crr* only, respectively (De Reuse and Danchin, 1988). In *B. subtilis*, the expression of sucrose-specific EII of PTS is induced by sucrose (Fouet et al., 1989).

The expression of *ptsHI* operon of *E. coli* is under very complex regulation. It is controlled by two contradictory regulatory mechanisms, the glucose-mediated and the CAP-cAMP-mediated activations (De Reuse et al., 1992; Ryu and Garges, 1994) which may involve three, possibly five, promoters (Postma et al., 1993; Ryu and Garges, 1994). Three of these

Figure 2. Schematic depiction of representative PTS permeases and their energy-coupling proteins fround in enteric bacteria (A to D and F), *B. subtilis* (E), and *R. capsulatus* (G) (Saier and Reizer, 1993).


Out
In
(A)
Mannitol
Mannitol-1-P
IIC Mtl
IIB
IIA
P
P
(B)
Glucose
P
Glucose-6-P
IIC Glc
IIB
IIA
P
(C)
Cellobiose
Cellobiose-6-P
IIC Cel
IIB
IIA
H
I
PEP
P
P
P
P
P
P
IIC Man
IID Man
IIB
IIA
(D)
Mannose
Mannose-6-P
P
P
IIC Fru
IID Fru
IIB
IIA
(E)
Fructose
Fructose-6-P
[DTP]
IIC Fru
IIB
IIB'
IIA
M
H
P
P
P
(F)
Fructose
Fructose-1-P
[MTP]
IIC Fru
IIB
IIB'
IIA
H
I
P
P
P
P
(G)
Fructose
Fructose-1-P

promoters (P0, P1, and Px) are located upstream from *ptsH*, separated by about 100 bp. A fourth (P2), the major promoter for *crr*, is located within the 3'-terminal end of *ptsI*. A fifth transcription start site, possibly involved in antitermination, is located downstream from *ptsH* and is directed against P0 and P1. In the P0-Px-P1 promoter region, there are two sequences showing similarity with the consensus of CAP-binding sites, one of which, CAPa, is located near P0, and the other, CAPb, is located in the -35 region of P0. P0 is the major promoter in the presence of the CAP-cAMP complex and P1 is the major promoter in its absence. In the absence of CAP, P0 is occluded. In the presence of CAP, binding of CAP to the P0 site is increased by P1. P2 is not activated by the CAP-cAMP complex. Most recently, Ryu and Garges (1994) showed that P0 and P1 exhibit a switching mechanism *in vitro*, where, depending upon the presence or absence of CAP-cAMP, transcription is initiated from different start sites (P0a and P0b, P1a and P1b). *In vivo* P0a is the major start site for P0, and P0b, a site 3 bases upstream from P0a, is never detected, although P0b is active *in vitro* when linear DNA is used and CAP-cAMP is present. There is no P0 expression when CAP or cAMP are absent in the cell unless glucose is present. *In vivo*, most transcription from P1 is from P1a unless glucose is present. Glucose causes a switch to P1b, a site 7 bases downstream from P1a. But in both events, the overall expression of P1a and P1b do not change much, meaning that there is very little CAP-cAMP activation of this promoter *in vivo*. Apparently, some barrier to P1b initiation can be relieved by the addition of glucose. Px is recognized by σ^{32} RNA polymerase, which is absent unless CAP-cAMP or glucose are present. All of these results indicate that the two different regulatory mechanisms (one through CAP-cAMP, the other through glucose) are working together for fine control of *ptsHI* operon.

Transcription antitermination is another way to regulate the expression of PTS proteins. One such example is the expression of *bgl* operon in decryptified *E. coli* strains. A model on this antitermination mechanism has been proposed (Postma et al., 1993). This operon is preceded by a constitutive promoter and consists of three genes, *bglG, bglF*, and *bglB*. These genes encode antiterminator, EII^{Bgl}, and phospho-ß-glucosidase, respectively. The first gene, *bglG*, is flanked on both sides by two terminators t1 and t2. The transcription of the operon is terminated by t1 and t2 in the absence of inducer, ß-glucosides. The antiterminator, BglG, can be phosphorylated in a reversible reaction by $\text{P-EII}^{\text{Bgl}}$. The phosphorylated BglG is inactive. In the presence of ß-glucosides, the uptake of ß-glucosides causes the phosphoryl group on $\text{P-EII}^{\text{Bgl}}$ to preferentially transfer to the substrate. This drain causes the dephosphorylation of P-BglG and subsequently become an active antiterminator. The active BglG binds to specific RNA sequences located upstream of t1 and t2, prevents termination, and makes the transcription of *bgl* operon possible. In *B. subtilis*, expression of *sacPA* operon which encodes a EIIBC^{scr} (*sacP*) and a sucrase (*sacA*) is also controlled by a proposed antiterminator, SacT. This antiterminator is homologous to the *E. coli* BglG, but requires EI and HPr for activity (Arnaud et al., 1992).

Many other mechanisms exist to regulate the expression of PTS proteins which include regulation through classical regulator-operator pairs, such as the expression of *nag* genes from *E. coli* and *K. pneumoniae* (Postma et al., 1993).

Metabolic Regulation of Non-PTS Sugars by PTS

In addition to the primary function in sugar transport, the PTS has an important role in the regulation of metabolism of non-PTS sugars. It has been observed in enteric bacteria that mutants defective in EI and HPr are not only unable to utilize PTS carbohydrates as the sole source of carbon for growth, but also unable to grow on a number of non-PTS sugars, including lactose, maltose, melibiose, glycerol, rhamnose, xylose, and Krebs cycle intermediates. This defect can be relieved by the addition of cAMP or by the introduction of a *crr* mutation (Postma et al., 1993). A model for this type of PTS-mediated regulation in enteric bacteria has been proposed (Postma et al., 1993). The central regulatory molecule is EIIA^{Glc}, the product of *crr*. EIIA^{Glc} can exist in two states: phosphorylated EIIA^{Glc} (P-EIIA^{Glc}) and nonphosphorylated EIIA^{Glc}. The phosphorylation state of EIIA^{Glc} is determined by the balance between phosphorylation via P-HPr and dephosphorylation via EIICB^{Glc} in the presence of its substrate. EII^{Glc} and P-EII^{Glc} interact with and regulate different proteins. EII^{Glc} binds to and inhibits several enzymes essential in carbohydrate metabolism, for example, the lactose and melibiose carriers. But the binding of EII^{Glc} occurs only when a substrate of the target protein is present. The direct result of binding is inhibition of uptake and of subsequent metabolism of these carbon sources. P-EIIA^{Glc} is an activator of adenylate cyclase, the enzyme for the synthesis of cAMP. cAMP plays a central role in gene expression in enteric bacteria. Together with CRP, cAMP is involved in the (generally) positive regulation of a number of catabolic genes. The inhibition of transport and metabolism of non-PTS carbohydrates can be brought about by any PTS carbohydrate since all PTS carbohydrates can dephosphorylate P-EIIA^{Glc}, either directly via EIICB^{Glc} (if the PTS carbohydrate is glucose) or indirectly (transport and phosphorylation of all other PTS sugars via their respective EII complexes result in dephosphorylation of P-HPr). Since the phosphorylation of EIIA^{Glc} by P-HPr is a reversible process, the dephosphorylation of P-HPr by other EII complexes will dephosphorylate P-EIIA^{Glc}. Recently, van der Vlag et al. (1994) demonstrated a quantitative relationship between the amount of EIIA^{Glc} and the regulation of glycerol and maltose metabolism in *S. typhimurium*. The inhibition of glycerol and maltose uptake by the addition of a PTS sugar was sigmoidally related to the amount of $EIIA^{Glc}$. For complete inhibition of glycerol uptake, a minimal ratio of about 3.6 mol of $EIIA^{Glc}$ to 1 mol of glycerol kinase was required while a ratio of about 18 mol of $EIIA^{Glc}$ to 1 mol of MalK protein (part of maltose transport system and the target of $EIIA^{Glc}$) was required for complete inhibition of maltose uptake. Varying the level of $EIIA^{Glc}$ did not affect the growth rate on glycerol, the rate of glycerol uptake, or the synthesis of glycerol kinase. In contrast, the growth rate on maltose, the rate of maltose uptake, and the synthesis of maltose-binding protein increased two- to fivefold with increasing levels of $EIIA^{Glc}$. The synthesis of MalK protein was not affected by varying the levels of $EIIA^{Glc}$.

Many Gram positive organisms also show a distinct "glucose effect" (Meadow et al., 1990). Several mechanisms exist whereby expression of catabolic genes can be regulated in Gram positive bacteria. In *B. subtilis*, catabolite-repressible genes are controlled by multiple mechanisms rather than by a single global regulatory system, and these mechanisms are known to differ from those operating in enteric bacteria (Stewart, 1993). Deutscher et al. (1994) has demonstrated that phosphorylation of HPr at Ser^{46} by the ATP-dependent kinase is directly or indirectly involved in catabolite repression in *B. subtilis*. After Ser^{46} in HPr of a *B. subtlis* strain was replaced with a nonphosphorylatable alanyl residue, the catabolite repressive effects of glucose on gluconate kinase, glucitol dehydrogenase, and the mannitol-specific catabolic enzymes were abolished, but this mutation had no effect on the uptake either of the PTS substrates, glucose and mannitol, or of the non-PTS substrates, gluconate and glucitol. A comparable result was obtained in *S. salivarius* (Vadeboncoeur et al., 1994). The replacement of Met^{48} in HPr with a valine resulted in the inhibition of phosphorylation of HPr on Ser^{46} by the ATP-dependent kinase but did not prevent phosphorylation of HPr by enzyme I or the phosphorylation of EII complexes by HPr (His-P). However, the mutant had pleiotropic properties, including reduced growth on non-PTS sugars. In *Lactobacillus brevis*, glucose and lactose are transported via a proton symport. The activities of glucose and lactose permeases in this organism are also regulated by the phosphorylation of HPr on Ser^{46} (Ye et al., 1994a).

Other Functions of PTS

PTS also has many other functions, including regulation of gluconeogenesis and other processes, interaction with acetate kinase, nitrogen regulation (Postma et al., 1993). The PTS is also a signal transduction system through which bacteria respond by positive chemotaxis to the presence of PTS carbohydrates. In PTS-dependent chemotaxis, stimulation corresponds to uptake and phosphorylation of a substrate through an EII. No methyl-accepting chemotaxis proteins (CheA, CheY, CheW, and CheZ) are involved in this process, since mutants lacking the methyl-accepting chemotaxis proteins retained normal PTS-dependent chemotaxis. Cells lacking a specific EII do not show chemotaxis toward the corresponding substrate. In contrast, mutants lacking EI and HPr do not show a chemotaxis response to any PTS carbohydrate, even when the corresponding EIIs are present.

CHAPTER 3

ISOLATION OF THERMOPHILIC CELLULOLYTIC *BACILLUS* STRAINS

Most thermophilic bacteria are generally isolated from hot springs, solfatatas or geothermally heated soils. But thermophilic species of *Bacillus* may be isolated from a wide range of both thermophilic and mesophilic environments. They comprise the first group of thermophilic microorganisms to be described and characterized in detail. The first published reference is attributed to Miquel in 1918 who described an organism growing at 70°C and that had been isolated from the river Seine in Paris (Sharp et al., 1992).

The genus *Bacillus* is one of the most diverse and commercially useful groups of microorganisms. Our understanding of the genetics and physiology of one of representatives, *B. subtilis*, is second only to that of *E. coli* (Harwood, 1989). Representatives of this genus are widely distributed in soils, air, and water where they are involved in a range of chemical transformations. The metabolic diversity and low reported incidence of pathogenicity have led to representatives of this group of organisms being used in a wide range of industrial processes. Strains of *Bacillus* are used for the production of four main types of products: (1) enzymes, (2) antibiotics, (3) fine biochemicals including flavor enhancers and food supplements, and (4) insecticides. It is the enzyme market that currently accounts for the main commercial importance of members of this genus. In particular they are the major source of hydrolyzing enzymes for the food and detergent markets.

Complete biodegradation of crystalline cellulose requires the concerted action of at least three types of enzymes which include cellobiase (see Review of Literature). Various methods have been developed for the screening of microorganisms which are able to degrade crystalline cellulose (Kluepfel, 1988).

The objective of this study was to isolate and identify thermophilic *Bacillus* strains which have cellulolytic activities and other hydrolase activities.

Materials and Methods

Sources of Samples

102 natural samples were screened for the isolation of thermophilic cellulolytic *Bacillus* strains. These samples included animal manures, composts, foods, pond water, sea water, soils, etc.

Isolation of Thermophilic *Bacillus* Strains with Putative Cellulolytic Activity

Approximately, 5 grams of environmental samples were combined with 5 ml of nutrient broth (3 g beef extract, 5 g peptone per liter) and mixed vigorously by votexing for approximately 5 minutes. After a brief period to allow the settling of large particles, 1 ml of this suspension was combined with 1 ml of 100% ethanol and shaken gently for 1 hour at room temperature (22°C) to enrich for endospores by inactivating vegetative cells (Koransky et al., 1978). Dilutions of treated samples were spread on nutrient agar plates and incubated overnight at 55°C or 65°C. Single colonies were transferred to master plates (nutrient agar) and to nutrient agar plates containing 0.25% powdered cellulose (Solka-Floc SW40; James River Corp., Berlin, N.H.). Putative cellulolytic isolates were identified by the presence of a clear zone or a partially cleared halo after 4 to 7 days of incubation at the temperatures from which they were isolated.

Testing for Hydrolases and Taxonomic Traits

The presence of complete cellulase activity was tested by using nutrient agar plates containing 0.25% cellulose as described above. Endoglucanase activity was tested by the Congo red method with nutrient agar plates containing 0.2% carboxymethyl cellulose (CMC) (Kluepfel, 1988). Putative cellobiohydrolase and ß-glucosidase activities were tested on nutrient agar plates containing 10 mg of 4-methylumbelliferyl-ß-D-cellobiopyranoside (MUC) or 4-methylumbelliferyl-ß-D glucopyranoside (MUG) per liter (Wood and Bhat, 1988). Mannosidase, arabinosidase, and xylosidase activities were screened in a similar manner with the respective umbelliferyl derivatives. The presence of pectin-degrading enzymes was tested by a modification of the method described by Brown et al. (1991) in which sodium polypectate was used instead of alginate. Taxonomic tests were performed as described in Bergey's manual (Claus and Berkeley, 1986).

Results

Thermophiles were abundant in all samples screened (2,000 to 1,500,000 CFUs per gram sample). All strains which were highly cellulolytic (clear zones on cellulose agar) were provisionally identified as actinomycete based on cell morphology and growth habit. However, many other isolates were surrounded by a partially cleared halo and were presumed to contain incomplete cellulase digestion systems.

A total of 168 representative clones which produced partially cleared halos on cellulose plates were selected for further study (112 were from the 55°C selection and 56 were from 65°C). All strains were aerobic, endospore-forming, Gram positive rods. Among the 112 strains from the 55°C selection, only 28 could grow at 65°C. In contrast, strains selected at 65°C also grew at 55°C and 75°C. More detailed taxonomic studies were conducted with 2 representative strains from the 55°C selection and 5 from the 65°C selection (Table 3).

Closely related organisms may have very similar proteins. In a separate project (Lai and Ingram unpublished data), a plasmid containing the genomic DNA of strain XL-65-6 was cloned and partially sequenced. A search of GenBank data base revealed that the deduced amino acid sequence of this insert

Table 3. Differential Properties of 55°C and 65°C Isolates

Properties	55°C Isolates		65°C Isolates				
	XL-55-60	XL-55-84	XL-65-1	XL-65-6	XL-65-11	XL-65-25	XL-65-56
Growth at							
30°C	+	+	-	-	-	-	-
40°C	+	+	+	+	+	+	+
55°C	+	+	+	+	+	+	+
65°C	-	-	+	+	+	+	+
75°C	-	-	+	+	+	+	+
Growth in							
2% NaCl	-	-	-	-	-	-	-
5% NaCl	-	-	-	-	-	-	-
7% NaCl	-	-	-	-	-	-	-
Growth at							
pH 5.7	+	+	+	+	+	+	+
pH 6.8	+	+	+	+	+	+	+
Acid from							
Glucose	+	+	+	+	+	+	+
Xylose	-	-	-	+	+	+	-
Hydrolysis of							
Casein	+	+	-	-	+	+	+
Gelatin	+	+	-	+	+	+	+
Starch	+	+	-	+	+	+	+
Catalase	+	+	+	+	+	+	+
pH in VP-broth	6.3	6.3	6.3	6.3	6.3	6.3	6.3

was highly homologous with the carboxyl terminal end of alanine racemase from *B. stearothermophilus, B. subtilis, E. coli*, and *S. typhimurium* (Table 4). It is interesting to note that the deduced amino acid sequence of this insert shares 91.4% identity and 96.2% similarity with the *B. stearothermophilus* enzyme, but shares relatively low homology with the enzymes from other organisms.

These organisms represent a rich source of enzymes for the depolymerization of carbohydrates (Table 5 and Table 6). All but 21 of the 168 isolates exhibited activity on one or more cellulolytic substrates (MUC, MUG, CMC), further validating the selection procedure. No CMCase activity was found in the isolates selected at 65°C or those selected at 55°C but able to grow at 65°C when tested at 65°C. More than half of the isolates which were selected at 55°C and unable to grow at 65°C exhibited endoglucanase activity. About one-third of the 65°C-growing isolates did, however, display weak CMCase activity when incubated at 55°C. All isolates contained multiple activities associated with the hydrolysis of cellulose and hemicellulose as well as other polymers, as illustrated by the representative strains (Table 5 and Table 6).

The profile of types of cellulase and hemicellulase activities in 65°C- and 55°C-growing strains are fairly different. Comparing the cellulase activities, MUG activity is the most abundant activity in 65°C-growing strains. The

Table 4. Comparison of amino acid sequences in carboxyl terminal portion of alanine racemases from XL-65-6 and other organisms

Organisms	% Identity / % Similarity[a] XL-65-6	Reference
Bacillus stearothermophilus	91.4 / 96.2	Tanizawa et al., 1988
Bacillus subtilis	60.2 / 75.8	Ferrari et al., 1985
Escherichia coli	40.7 / 61.1	Lobocka et al., 1992
Salmonella typhimurium	30.3 / 57.0	Galakatos et al., 1986

[a] The results were from the comparison of 186 amino acids at the carboxyl terminal end of XL-65-6 alanine racemase with corresponding region and number of amino acids of alanine racemases from other organisms.

Table 5. Enzymatic Activities Expressed by Isolates Capable of Growing at 65°C or Higher

Enzymes and Substrates[a]	No. Strains[b]	% Total	XL-65-1	XL-65-6	XL-65-11	XL-65-25	XL-65-56
Cellulases							
CMC[c]	24	28.6	+	−	−	−	+
MUC	41	48.8	+	+	+	+	+
MUG	59	70.2	+	+	+	+	+
CMC/MUC	11	13.1	+	−	−	−	+
CMC/MUG	20	23.8	+	−	−	−	+
MUC/MUG	38	45.2	+	+	+	+	+
CMC/MUC/MUG	7	8.3	+	−	−	−	+
Hemicellulases							
MUA	20	23.8	−	+	+	+	+
MUM	65	77.3	+	+	+	+	+
MUX	27	32.1	−	−	+	+	+
MUA/MUM	17	20.2	−	+	+	+	+
MUA/MUX	16	19.0	−	−	+	+	+
MUM/MUX	20	23.8	−	−	+	+	+
MUA/MUM/MUX	12	14.3	−	−	+	+	+
Other enzymes							
Milk	24	28.6	−	−	+	+	+
Pullulan	18	21.4	+	+	−	−	−
Starch	58	69.0	−	+	+	+	+
Sodium polypectate[d]	39	46.4	+	+	+	−	−

[a] CMC: carboxymethyl cellulose
MUC: 4-methylumbelliferyl-β-D-cellobiopyranoside
MUG: 4-methylumbelliferyl-β-D-glucopyranoside
MUA: 4-methylumbelliferyl-α-L-arabinofuranoside
MUM: 4-methylumbelliferyl-β-D-mannopyranoside

Table 5 -- contiued

MUX: 4-methylumbelliferyl-ß-D-xyloside

[b] A total of 84 strains were isolated and tested.

[c] Results were from tests at 55°C incubation. None were active at 65°C.

[d] No attempt was made to distinguish hydrolase from lyase activities.

Table 6. Enzymatic Activities Expressed by Isolates Capable of Growing at 55°C or Lower

Enzymes and Substrates[a]	No. Strains[b]	% Total	XL-55-60	XL-55-84
Cellulases				
CMC	58	69.0	+	+
MUC	68	80.9	+	+
MUG	64	76.2	+	+
CMC/MUC	49	58.3	+	+
CMC/MUG	47	56.0	+	+
MUC/MUG	63	75.0	+	+
CMC/MUC/MUG	46	54.8	+	+
Hemicellulases				
MUA	52	61.9	+	+
MUM	16	19.0	-	-
MUX	44	52.4	+	+
MUA/MUM	13	15.5	-	-
MUA/MUX	40	47.6	+	+
MUM/MUX	8	9.5	-	-
MUA/MUM/MUX	8	9.5	-	-
Other enzymes				
Milk	49	58.3	+	+
Pullulan	5	6.0	-	-
Starch	36	42.9	+	+
Sodium polypectate[c]	17	20.2	-	-

[a] The abreviations of substrates see Table 5

[b] A total of 84 strains were isolated and tested

[c] No attempt was made to distinguish hydrolase from lyase activities.

percentage of strains with this activity (70.2%) is predominantly higher than those with CMC or MUC activities (28.6% and 48.8%, respectively). Only 8.3% of the strains exhibited all three cellulase activities. In 55°C-growing strains, all three cellulase activities are fairly common. Approximately 80% of the strains exhibited MUC activity, 76% exhibited MUG activity, and 69% exhibited CMC activity. Over half of these strains contained all of the three activities.

For the three hemicellulase activities tested (arabinosidase, mannosidase, and xylosidase), these two groups of bacteria also exhibited different profiles. In 65°C-growing strains, MUM activity was the most common activity (77%) with MUA and MUX activities far less common (23% and 32%, respectively). In contrast, MUM activity was the least common activity in 55°C-growing stains. Only 19% of the strains had this activity, while more than half of the strains had MUA and MUX activities (61% and 52%, respectively). Strains which exhibited all three hemicellulase activities were rare in both groups of strains, with 14% for 65°C-growing strains and 9% for 55°C-growing strains.

Other hydrolytic activities are also common in all the isolates. Approximately 69% of the 65°C-growing strains exhibited amylase activity and 58% of the 55°C-growing strains had caseinase activity.

Strain XL-65-6, which was isolated from rotting wood by Lake Alice (Gainesville, Florida), was selected for further study on cellobiose utilization at a molecular level since it had high activities on both MUC and MUG, as well as on MUA and MUM (Table 5).

Discussion

The objective of this study was to isolate thermophilic cellulolytic strains of *Bacillus*. The isolation method used, combining ethanol pre-treatment, high incubation temperature, and nutrient agar medium, appeared to be fairly selective for thermophilic *Bacillus* species from various habitats. All colonies isolated were either *Bacillus* species or actinomycetes. *Bacillus* species constituted the majority of these isolates. Since most of actinomycete also form spores and many are thermophiles (Brock and Madigan, 1991), it was not surprising that members of this group were also recovered with *Bacillus* species although the number of actinomycete was much lower.

Representative strains were classified according to Bergey's Manual (Claus and Berkeley, 1986). All strains had the distinctive characteristics of *Bacillus* species - aerobic, endospore-forming, Gram positive rods.

The *Bacillus* species which can grow at 65°C or higher include *B. acidocaldarius*, *B. schlegelii*, and *B. stearothermophilus*. But *B. acidocaldarius* can grow only under acidic condition (below pH 6) while the 65°C isolates from this study grew well at pH 6.8 (Table 3). *B. schlegelii* does not produce acid from glucose and xylose, and lacks the ability to hydrolyze gelatin and starch. Most of the 65°C isolates produced acid by fermenting these two sugars and hydrolyzed gelatin and starch. In contrast, most of the traits of the isolates fit those of *B. stearothermophilus*. Furthermore, one 65°C isolate, XL-65-6, had very similar alanine racemase to that from *B. stearothermophilus*. Thus, these 65°C isolates were identified as *B. stearothermophilus*.

Among the *Bacillus* species, only *B. coagulans* and *B. licheniformis* strains are capable of growing at temperatures ranging from 30°C to 55°C. *B. licheniformis* are capable of growing at high NaCl concentration (up to 7%), while *B. coagulans* cannot grow at 5% of NaCl. Both strain XL-50-60 and XL-50-84 grew at 30°C and 55°C, but not at 65°C, and neither grow in the presence of 2% NaCl. Thus, the strains which were isolated at 55°C and unable to grow at 65°C are most similar to *B. coagulans*. It is likely that the strains which were isolated at 55°C and able to grow at 65°C were strains of *B. stearothermophilus*.

Many isolates did not exhibit all properties listed in Bergey's Manual. For example, 11% - 89% of *B. stearothermophilus* strains can grow at 5% NaCl. however, all of our isolates failed to grow in 2% NaCl. As mentioned in the Literature Review section, both *B. coagulans* and *B. stearothermophilus* are highly heterogeneous species. The positive traits in Bergey's Manual indicate only that more than 90% of strains are positive. The occurrence of some traits which are different from the standard strains in Bergey's Manual is thus not surprising.

The activities on MUC and MUG, as well as those of MUA and MUM appear to be linked. Among *B. stearothermophilus* strains, 45.2% had both MUC and MUG activities. This percentage is very close to that showing MUC activity alone (48.8%). The percentage of strains with CMC and MUC or MUG activity was different from those with only one activity. The percentage of *B. coagulans* strains expressing both MUC and MUG activities was 75.0%, similar to MUG activity alone (76.2%). Those containing CMC activity were also quite different. There are at least three possible explanations for these results. First, genetic linkage, where in *B. stearothermophilus* and *B. coagulans* MUG and MUC activity are linked. Second, ambiguity of the substrates, MUC and MUG. The chromogenic substrate MUC has often been used to measure cellobiohydrolase activity, but potential confounding activities from ß-glucosidases are well known (Coughlan, 1988). Thus, both substrates may be cleaved by a single enzyme. Third, multiple activities, where one enzyme truly has both exoglucanase and ß-glucosidase activity. A comparable situation may occur for MUA and MUM activity.

CHAPTER 4

CLONING AND SEQUENCING OF CELLOBIOSE-SPECIFIC PERMEASE OF PHOSPHOENOLPYRUVATE-DEPENDENT PHOSPHOTRANSFERASE SYSTEM FROM *BACILLUS STEAROTHERMOPHILUS*

Despite the potential abundance of cellobiose in the environment, comparatively little is known about the uptake systems which initiate metabolism in bacteria. The best characterized *cel* operon is cryptic and must be activated in laboratory strains of *E. coli* (Parker and Hall, 1990). Functional genes for cellobiose utilization have been reported in some recent isolates of *E. coli* from human and animal manures (Hall and Faunce, 1987). Cellobiose is also metabolized by other enteric bacteria such as *Klebsiella* sp. (Al-Zaag, 1989) and *Erwinia* sp. (El Hassouni et al., 1992). Considering the similarities between these enteric genera, it is likely that all use a cellobiose PTS to initiate metabolism. Many Gram positive and Gram negative bacteria have been reported to contain cell-associated ß-glucosidases (Coughlan and Mayer, 1992). Gram positive bacteria such as *Cellulomonas uda* and *C. favigena* contain cellobiose phosphorylases which are presumed to be involved in intracellular cellobiose metabolism (Coughlan and Mayer, 1992).

The objective of this study was the cloning and sequencing of *Bacillus stearothermophilus* genes which confer the ability to hydrolyze the cellobiose model substrates, 4-methylumbelliferyl-ß-D-cellobiopyranoside (MUC) and 4-methylumbelliferyl-ß-D-glucopyranoside (MUG).

Materials and Methods

Bacterial Strains, Plasmids, and Growth Conditions

Bacterial strains and plasmids used in this study are listed in Table 7. Strains of *B. stearothermophilus* were grown in Difco Nutrient Broth (or agar) at 65°C. Strains of *E. coli* were grown at 37°C in Luria broth (or agar) (Atlas and Parks, 1993). Ampicillin (50 µg/ml) and tetracycline (10 µg/ml) were added to media after sterilization as appropriate for selection of recombinant *E. coli*.

DNA Manipulation

Procedures used for the preparation of *E. coli* plasmids, assembly of recombinant DNA, and the transformation of *E. coli* were as described previously (Sambrook et al., 1989). Digestions with restriction enzymes were carried out as recommended by the manufacturers. Genomic DNA was isolated from *B. stearothermophilus* strain XL-65-6 by the standard method (Cutting and Vander Horn, 1990).

Cloning of *B. stearothermophilus cel* Operon

Initial testing indicated that DNA isolated from XL-65-6 was not efficiently digested by restriction enzyme *Sau*3AI but was digested into small fragments by *Mbo*I. Poor digestion by

Table 7. Strains and plasmids used in this study

Strain or plasmid	Genetic characteristics	Sources
Bacteria		
B. stearothermophilus	Prototroph	this study
E. coli		
DH5α	F^- *ΔlacZM15 deoR recA1 endA1 supE44 hsdR17*(r_K^-, m_K^+) λ^- *thi-1 gyrA96 relA1*	BRL^a
XL1-Blue	F'::Tn*10* (tet^r)*lacI^{q}Δ(lacZ)M15 recA1 endA1 gyrA96 thi hsdR17*(r_K^-, m_K^+) *supE44 relA1*	Stratagene^b
MM6	*lacI22 dctB3 ptsI2 relA1 thi-1 spoT1*	CGSC^c
Plasmids		
pUC18	*bla lacI'Z'*	BRL^a
pBluescript II KS$^-$	*bla lacZ*	Stratagene^b
pBluescript SK$^-$	*bla lacZ*	Stratagene^b
pLOI902	pUC18 containing *cel* operon	This study
pLOI905	pUC18 containing *cel* operon in opposite orientation	This study

[a] Bethesda Research Laboratories, Gaithersburg, MD
[b] Stratagene, La Jolla, CA
[c] *E. coli* Genetic Stock Center, Yale University, New Haven, CT

*Sau*3AI is presumed to result from the presence of 5-methylcytosine which is tolerated by *Mbo*I (Roberts and Macelis, 1992). Fragments of 4-6 kilobase pairs (kbp) were isolated from partial *Mbo*I digestions of chromosomal DNA by agarose gel electrophoresis. These were ligated into the *Bam*HI site of pUC18 and transformed into *E. coli* strain DH5α. Plasmid DNA was isolated from the pooled colonies (approximately 3000 clones) and served as an amplified library. Secondary transformants of DH5α were screened for MUC and MUG activity. Clones containing both activities were subsequently found to contain the *cel* operon.

Southern hybridization

Samples of genomic DNA from *B. stearothermophilus* XL-65-6 were digested with five restriction enzymes (*Bss*HII, *Eco*RI, *Hind*III, *Pst*I, and *Sty*I), respectively. After agarose gel electrophoresis, fragments were transferred to Zeta-Probe GT membranes (Bio-Rad Laboratories, Richmond, CA) and probed with a 1.1 kbp *Sty*I fragment from within the *B. stearothermophilus cel* operon. This probe was labeled with digitoxigenin (Genius System, Boehringer Mannheim Biochemicals, Indianapolis, IN) and used as recommended by the manufacturer. Genomic DNA from *E. coli* DH5α was digested with *Pst*I and served as a control.

DNA sequencing and sequence analysis

The 5 kbp fragment in pLOI902 was subcloned into pBluescript II KS$^-$ and pBluescript SK$^-$ prior to sequencing. Two series of nested deletions were generated by using the

Erase-a-Base System (Promega Corp., Madison, WI) to allow sequencing of the entire fragment in both directions. The Magic Miniprep DNA Purification System (Promega Corp., Madison, WI) was used to prepare double-stranded plasmid DNA. Sequencing was performed by the dideoxynucleotide-chain termination method of Sanger et al. (1977) using fluorescent-labeled primers (forward, 5'-CACGACGTTGTAAAACGAC-3'; forward, 5'-GGTTTTCCCAGTCACGACG-3'; reverse, 5'-ATAACAATTTCACACAGGA-3') (LI-COR, Inc., Lincoln, NE) and the Sequenase 7-deaza-dGTP DNA Sequencing kit (United States Biochemical Corp., Cleveland, OH). Regions of compression were resolved by using either dITP (United States Biochemical Corp., Cleveland, OH) or the TaqTrack Deaza Sequencing System (Promega Corp., Madison, WI). Sequencing was carried out on a LI-COR model 4000 DNA sequencer and image analyzer using 7% acrylamide gels (Long Ranger Gel Concentrate, AT Biochem, Malvern, PA).

Nucleotide and deduced amino acid sequences were manipulated and analyzed with the Genetics Computer Group Sequence Analysis Software Package (University of Wisconsin, Genetics Computer Group, Madison, WI) (Devereux et al., 1984) and the National Center for Biotechnology Information using the BLAST network service (Benson et al., 1993). Amino acid sequences were aligned by using the Clustal V program (Higgins et al., 1992). The nucleotide sequence data has been submitted to GenBank and assigned the accession number U07818.

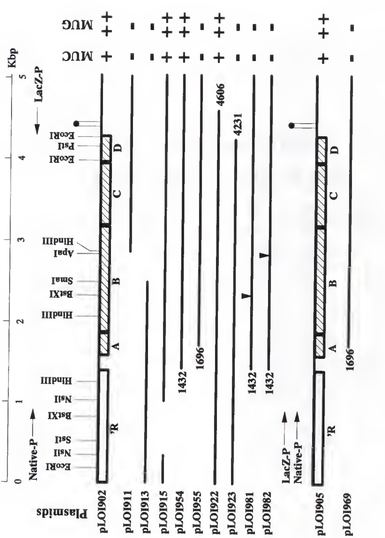
Results

Cloning of Genes Encoding MUC and MUG Activities (*cel* Operon)

The plasmid library was screened for activities hydrolyzing MUC and MUG after transformation into DH5α. Approximately 50 positive clones were identified from the 50,000 colonies tested. Those with weak activities were unstable. Nineteen clones were found which exhibited high activities for both substrates. These activities were present together in all clones. Digestions with restriction enzymes indicated that all active clones contained the same 5 kbp DNA fragment and were siblings. One of these, denoted pLOI902, was selected for further study.

A restriction map of pLOI902 is shown in Figure 3. The minimal coding region for the activities hydrolyzing MUC and MUG was localized by testing subclones. Subclones pLOI911, pLOI913, and pLOI915 were constructed by deleting *Hind*III, *Sma*I, and *Nsi*I fragments, respectively. All but pLOI915 were inactive. This subclone retained both activities. Subclones pLOI922, pLOI923, pLOI954, and pLOI955 were constructed by deleting terminal regions with exonuclease III. MUC and MUG activities remained after the deletion of ca. 400 bp adjacent to the *lac* promoter (pLOI922) or 1,500 bp from the opposite end (pLOI954). However, further deletion of 375 bp from pLOI922 or 264 bp from pLOI954 resulted in the concomitant loss of both activities (pLOI923 and pLOI955, respectively). Thus a rather large region of DNA (approximately 3.2 kbp)

Figure 3. Restriction map of the *B. stearothermophilus* DNA fragment contained in pLOI902. Subclones pLOI902, pLOI905, pLOI911, pLOI913, pLOI915, and pLOI969 are derivatives of pUC18. Subclones pLOI922 and pLOI923 are derivatives of pBluescript II KS$^-$. Subclones pLOI954 and pLOI955 are derivatives of pBluescript SK$^-$. Genes of *cel* operon are labled *A, B, C,* and *D*. The incomplete regulatory gene is labeled *'R*. The proposed terminator is marked as a solid ball. Approximate positions of frame-shift mutations are indicated by triangles in pLOI981 and pLOI982. These mutations were introduced by Klenow treatment and religation after digestion with *Bst*XI and *Apa*I, respectively. Derivates of pUC18 are aligned beneath restriction sites used for construction; others are aligned beneath their respective sequence positions. The activities of MUC and MUG are indicated at the right.


Plasmids
pLOI902
pLOI911
pLOI913
pLOI915
pLOI954
pLOI955
pLOI922
pLOI923
pLOI981
pLOI982
pLOI905
pLOI969
0
1
2
3
4
5
Kbp
Native-P
LacZ-P
EcoRI
NsiI
SstI
BstXI
NsiI
HindIII
HindIII
BstXI
SmaI
ApaI
HindIII
EcoRI
PstI
EcoRI
'R
A
B
C
D
1432
1696
4606
4231
MUC
MUG
+
++
-

appeared to be required for the expression of MUC and MUG activities in *E. coli* DH5α. Both activities were expressed or lost concurrently in all subclones. Active clones consistently hydrolyzed MUG more rapidly than MUC.

Experiments were conducted to determine the direction of transcription of the genes encoding MUC and MUG activities in pLOI902. The presence or absence of inducer (isopropyl-thio-ß-D galactopyranoside or IPTG) did not alter the level of expression in XL1-blue. Attempts to subclone this fragment into plasmid pUC18 in the opposite orientation were unsuccessful using DH5α as a host, but clones were readily obtained using XL1-Blue. The resulting plasmid (pLOI905) was lethal when transformed back into DH5α. The growth of XL1-Blue harboring pLOI905 was strongly inhibited by the presence of inducer (IPTG), consistent with partial control by the resident *lacI*^{q} in the F' element. Thus, in the original clone, pLOI902, the direction of transcription for the *cel* operon is opposite to that of the *lac* promoter.

Analysis of DNA Sequence

The entire 5,026 bp DNA fragment in pLOI902 was sequenced in both directions (Fig. 3 and 4). Four complete open reading frames were found in the region required for the two activities, bordered by incomplete genes at both ends. All appear to be transcribed in the same direction. The four central genes appear to be part of an operon (bases 1541 through 4306) with a potential promoter at the 5' end and a

Figure 4. Nucleotide sequence of the 5 kbp DNA fragment from *B. stearothermophilus* contained in pLOI902. The deduced amino acid sequences for the *cel* operon and the incomplete regulatory gene are listed below the first nucleotide of the corresponding codon. A putative promoter region for the *cel* operon is underlined with the -35 and -10 regions labeled. The sequence for the 3' terminus of 16s rRNA from *B. stearothermophilus* (Van Charldorp et al., 1981) is shown above the potential ribosomal binding region (underlined and labeled RB) from the *cel* genes. Unmatched bases are in lower case. Genes within the *cel* operon are marked at their start codons. Stop codons are indicated by asterisks. The proposed terminator is marked with inverted arrows which correspond to a proposed stem-loop structure.

```
   1 TCCCGACGTCCTTGAGTTGACCGAAGCGATGGTCGTCTATGCGGAGAAAAAACTTGGGCG
      P  D  V  L  E  L  T  E  A  M  V  V  Y  A  E  K  K  L  G  R
  61 CCGCCTCGCCGAGAAGGTGATGTATGCGCTCGCCATGCACATTCAAACGGCGATCAACCG
      R  L  A  E  K  V  M  Y  A  L  A  M  H  I  Q  T  A  I  N  R
 121 CCTGCGGGCCGGGACGATCGTTTCCCATCCGAAACTGAATGAAGTTCGCGCCGCCTATAA
      L  R  A  G  T  I  V  S  H  P  K  L  N  E  V  R  A  A  Y  K
 181 GCAGGAATTCGCCGTCGCGCTCGACTGCCTTCAGCTGATGGAAGAAAGGACGAATATCGA
      Q  E  F  A  V  A  L  D  C  L  Q  L  M  E  E  R  T  N  I  D
 241 CTTCCCGATCGATGAAGCGGGATTTTTGACGATGTTTTTTGCGTTCCACGAGGAGCAAGC
      F  P  I  D  E  A  G  F  L  T  M  F  F  A  F  H  E  E  Q  A
 301 GGAAGAGCGGGAGGAACGGGTGGCGATCGTGGTTGTGATGCATGGCAACGGTGTGGCTTC
      E  E  R  E  E  R  V  A  I  V  V  V  M  H  G  N  G  V  A  S
 361 GGCGATGGCGGACGTCGTGAATCAGCTGTTGGCCGCCGCGTGTGTGCATGCGGTCGATAT
      A  M  A  D  V  V  N  Q  L  L  A  A  A  C  V  H  A  V  D  M
 421 GCCGCTTGATGCCGACCCGAAACGAATTTACGAGCAGGTGAAAGCGGTGCTACAGCCCGT
      P  L  D  A  D  P  K  R  I  Y  E  Q  V  K  A  V  L  Q  P  V
 481 CGCATCGAAAAAAGGGGCGCTGTTGCTTGTTGACATGGGATCGCTCGTGTCGTTTTCGAA
      A  S  K  K  G  A  L  L  L  V  D  M  G  S  L  V  S  F  S  N
 541 CTTTTTGGAAAAAGAGCTCGCTGTTCCGGTGCGGGTGATTTCCGCTGCCAGCACGCCGCA
      F  L  E  K  E  L  A  V  P  V  R  V  I  S  A  A  S  T  P  H
 601 CGTGCTCGAAGCGGCCCGCAAAGCGATGCTCGGCTATGCGCTTCAGGAAATTTACGAAGA
      V  L  E  A  A  R  K  A  M  L  G  Y  A  L  Q  E  I  Y  E  E
 661 AGTGAAAGCCGCAGCGCCGTTTTACATTCGCGGGCCGCTTTGGGAGGAGGAGGCTGCGGA
      V  K  A  A  A  P  F  Y  I  R  G  P  L  W  E  E  E  A  A  E
 721 GCAGGACATGCTGGCGATCGTCACCGCCTGTTTTACTGGAAAAGGAAGCGCGCTGGCGCT
      Q  D  M  L  A  I  V  T  A  C  F  T  G  K  G  S  A  L  A  L
 781 AAAGCATATATTGGAAACGTATTTGCAGCTCGATGAGCGCATGTGGCGCATTATTCCCAT
      K  H  I  L  E  T  Y  L  Q  L  D  E  R  M  W  R  I  I  P  I
 841 TCAAATGGCAGATGCGGAAGAAGCGCGGCAAACCTTGTCCAACGTCGCGAAACATTTCCG
      Q  M  A  D  A  E  E  A  R  Q  T  L  S  N  V  A  K  H  F  R
 901 CATCGTCTGTATCGTCAGCCATCTTTGCCTTGATGAGCGAATCCCGCACTTTTCTCTTGA
      I  V  C  I  V  S  H  L  C  L  D  E  R  I  P  H  F  S  L  E
 961 GGACGTGCTTAGCTTGAGGGCGATGAAAGAAATCCAAGCGTTAGCCGACGTTGAGGAAAT
      D  V  L  S  L  R  A  M  K  E  I  Q  A  L  A  D  V  E  E  I
1021 TCATATGCATATGGCTAGAGAGCTGACCAATCATCTTCGCCACCTCGCACCGGcGCGGGC
      H  M  H  M  A  R  E  L  T  N  H  L  R  H  L  A  P  A  R  A
1081 CATTCCCGCCATTCGCGCTGCGCTGGCGGCGATTGGTCGGGAACTCGGCCTTGAAGCGGA
      I  P  A  I  R  A  A  L  A  A  I  G  R  E  L  G  L  E  A  D
1141 TGGCCGCGATCTGGTGGGGCTTGTCTTTCATCTTTGCTGCTTGCTCGATCGCCTGCTGTC
      G  R  D  L  V  G  L  V  F  H  L  C  C  L  L  D  R  L  L  S
1201 CGGAGAGAGAACGAGCGGTGACCGAGGAGCAGCGAAAGCTTGCGGGCATGAAGATGAAGA
      G  E  R  T  S  G  D  R  G  A  A  K  A  C  G  H  E  D  E  D
1261 TGGCGCACTGTACAGGGCGGTCAAGGAGGGGTTGTTTCCGCTTGAACAGCAGTACGGTGT
      G  A  L  Y  R  A  V  K  E  G  L  F  P  L  E  Q  Q  Y  G  V
1321 CTGCATCGATGAAGATGAACTGTGTCATATCGTTCACTTTTTTCGCTCCCTGCAAGGAAA
      C  I  D  E  D  E  L  C  H  I  V  H  F  F  R  S  L  Q  G  K
1381 ACGAGATGGATAAGGTGCGATAATTGGAAACGCTTCCATTACACAGTTTCCCTCCGTTAA
      R  D  G  *
1441 CACACAACCGTTCCATCCGCTTCAGCGCTGTCTTTTCGGCGTTGGCATGGAATTTGCTTG
                               UUUCCUCC                 -35
1501 AACAATAATGAATCATTGCAAAAAAAGGAGGAGGAAGCGGATGAACATTTTGCTCATTTG
           -10                    RBS               M  N  I  L  L  I  C
                                                    CelA →
```

```
1561 CGCTGCCGGCATGTCGACGAGTTTGCTTGTGACGAAGATGAAGGAGGCGGCAAAGCAAAA
      A  A  G  M  S  T  S  L  L  V  T  K  M  K  E  A  A  K  Q  K
1621 AGGGATCGAGGCGAACATTTGGGCTGTGTCAGCTGATGAAGCGAAAAGTCATCTCGACCA
      G  I  E  A  N  I  W  A  V  S  A  D  E  A  K  S  H  L  D  Q
1681 GGCGGATGTTGTGCTGATTGGTCCGCAAATCCGTTATAAGCTCGCCGCCTTCAAAAAAGA
      A  D  V  V  L  I  G  P  Q  I  R  Y  K  L  A  A  F  K  K  E
1741 GGGAGAGGCGCGCGGCATTCCGGTTGACGTCATCAATCCAGCTGACTATGGGCGGGTCAA
      G  E  A  R  G  I  P  V  D  V  I  N  P  A  D  Y  G  R  V  N
                                              UUUccUCCaC
1801 CGGCGCCGGCGTGTTGGACTTTGCGCTGCGGTTAAAAAAATAAACAGGGGGTAGAAGCGG
      G  A  G  V  L  D  F  A  L  R  L  K  K  *     RBS
1861 ATGGACCGGTTTATTCGGATGTTGGAAGACCGCGTGATGCCTGTCGCCGGCAAGATTGCC
     M  D  R  F  I  R  M  L  E  D  R  V  M  P  V  A  G  K  I  A
     CelB →
1921 GAACAGCGCCATTTGCAGGCGATTCGTGACGGGATTATTTTGTCGATGCCATTGTTGATT
     E  Q  R  H  L  Q  A  I  R  D  G  I  I  L  S  M  P  L  L  I
1981 ATCGGGTCTTTATTTTTAATCGTTGGCTTTTTGCCGATTCCCGGTTACAACGAATGGATG
     I  G  S  L  F  L  I  V  G  F  L  P  I  P  G  Y  N  E  W  M
2041 GCGAAATGGTTCGGCGAGCATTGGCTTGATAAGCTTCTCTATCCGGTTGGAGCGACGTTC
     A  K  W  F  G  E  H  W  L  D  K  L  L  Y  P  V  G  A  T  F
2101 GATATTATGGCGCTTGTTGTCAGCTTCGGAGTCGCCTATCGGCTGGCGGAAAAGTATAAA
     D  I  M  A  L  V  V  S  F  G  V  A  Y  R  L  A  E  K  Y  K
2161 GTGGATGCGCTGTCCGCCGGCGCGATTTCACTGGCTGCTTTTTTGCTCGCAACCCCGTAT
     V  D  A  L  S  A  G  A  I  S  L  A  A  F  L  L  A  T  P  Y
2221 CAAGTGCCGTTCACGCCGGAAGGAGCGAAAGAAACCATTATGGTCAGCGGTGGCATCCCG
     Q  V  P  F  T  P  E  G  A  K  E  T  I  M  V  S  G  G  I  P
2281 GTGCAATGGGTCGGCAGCAAGGGGTTGTTTGTTGCCATGATTTTGGCGATTGTGTCAACC
     V  Q  W  V  G  S  K  G  L  F  V  A  M  I  L  A  I  V  S  T
2341 GAAATTTACCGGAAAATCATTCAAAAAAATATTGTCATTAAGCTGCCGGACGGGGTGCCG
     E  I  Y  R  K  I  I  Q  K  N  I  V  I  K  L  P  D  G  V  P
2401 CCTGCTGTGGCCCGCTCCTTTGTTGCTTTGATCCCGGGAGCCGCCGTTCTCGTCGTTGTC
     P  A  V  A  R  S  F  V  A  L  I  P  G  A  A  V  L  V  V  V
2461 TGGGTAGCCCGGCTGATTTTGGAAATGACACCGTTTGAAAGTTTCCATAACATTGTATCT
     W  V  A  R  L  I  L  E  M  T  P  F  E  S  F  H  N  I  V  S
2521 GTCCTCCTGAACAAACCGCTCAGTGTGCTCGGCGGCAGTGTATTTGGCGCCATTGTCGCC
     V  L  L  N  K  P  L  S  V  L  G  G  S  V  F  G  A  I  V  A
2581 GTGCTGCTTGTCCAGCTGCTATGGTCGACCGGACTGCACGGAGCGGCCATCGTCGGCGGG
     V  L  L  V  Q  L  L  W  S  T  G  L  H  G  A  A  I  V  G  G
2641 GTCATGGGGCCGATCTGGCTGTCGCTGATGGATGAAAATCGAATGGTGTTCCAGCAAAAT
     V  M  G  P  I  W  L  S  L  M  D  E  N  R  M  V  F  Q  Q  N
2701 CCGAATGCCGAACTGCCCAACGTCATTACGCAACAGTTTTTTGATCTTTGGATTTACATC
     P  N  A  E  L  P  N  V  I  T  Q  Q  F  F  D  L  W  I  Y  I
2761 GGCGGTTCAGGAGCGACATTGGCGTTGGCGTTGACGATGATGTTTCGGGCCCGCAGCCGG
     G  G  S  G  A  T  L  A  L  A  L  T  M  M  F  R  A  R  S  R
2821 CAGCTGAAAAGCTTAGGTCGGCTGGCGATCGCGCCCGGCATTTTCAATATTAATGAGCCG
     Q  L  K  S  L  G  R  L  A  I  A  P  G  I  F  N  I  N  E  P
2881 ATCACGTTCGGCATGCCGATCGTCATGAACCCATTGCTGATCATTCCGTTCATTCTCGTG
     I  T  F  G  M  P  I  V  M  N  P  L  L  I  I  P  F  I  L  V
2941 CCTGTCGTGCTTGTGGTTGTTTCCTACGCCGCGATGGCGACCGGGCTTGTCGCCAAACCA
     P  V  V  L  V  V  V  S  Y  A  A  M  A  T  G  L  V  A  K  P
3001 AGCGGGGTGGCCGTGCCATGGACGACACCGATCGTGATCAGCGGCTATTTAGCGACGGGA
     S  G  V  A  V  P  W  T  T  P  I  V  I  S  G  Y  L  A  T  G
```

```
3061 GGCAAAATTTCCGGGAGTATTTTGCAAATCGTTAACTTCTTCATCGCGTTTGCCATCTAC
     G  K  I  S  G  S  I  L  Q  I  V  N  F  F  I  A  F  A  I  Y
3121 TATCCATTTTTCTCGATTTGGGACAAACAAAAAGCGGCCGAAGAGCAGGCCGATCCAACG
     Y  P  F  F  S  I  W  D  K  Q  K  A  A  E  E  Q  A  D  P  T
                                       UUUCCUC
3181 ATTTCAAGCGGAGCGGGGGCAACGCACTCGCTGTAAAGGAGAGGAACGAAATGCCGCGCT
     I  S  S  G  A  G  A  T  H  S  L  *  RBS           M  P  R  Y
                                                       CelC →
3241 ATTGCATCGTCAACGCCGATGATTTCGGTTACTCGAAAGGGGTCAACTACGGGATTTTGG
       C  I  V  N  A  D  D  F  G  Y  S  K  G  V  N  Y  G  I  L  E
3301 AAGCGTTTCAGAACGGTGTCGTCACGTCGGCGACGCTGATGGCGAATATGCCAGCGGCCG
       A  F  Q  N  G  V  V  T  S  A  T  L  M  A  N  M  P  A  A  E
3361 AACACGCCGCCCGGCTGGCGAAGGACCATCCGGAACTCGGCGTTGGCATTCATTTTGTGC
       H  A  A  R  L  A  K  D  H  P  E  L  G  V  G  I  H  F  V  L
3421 TGACGTGCGGCCGGCCGCTGGCCGATGTTCCATCGCTGGTGAACGAGAATGGGGAGTTTC
       T  C  G  R  P  L  A  D  V  P  S  L  V  N  E  N  G  E  F  P
3481 CGCGGCGCGGGGAGGCGCTTGTCGGCGCTAGGCGCGGCGATATCGAGCGGGAGCTTTGCG
       R  R  G  E  A  L  V  G  A  R  R  G  D  I  E  R  E  L  C  A
3541 CCCAATTGGAGCGTTTTTTCTCGTTCGGGCTCACTCCGACGCATATTGACAGCCATCATC
       Q  L  E  R  F  F  S  F  G  L  T  P  T  H  I  D  S  H  H  H
3601 ACGTTCATGAGCACCCGAATGTGTTTCCGGTTGTGGAACAATTGGCCGAACGCTATCGGC
       V  H  E  H  P  N  V  F  P  V  V  E  Q  L  A  E  R  Y  R  L
3661 TGCCGATCCGCCCGGTGCGGACGGCACGGCCGCATCGGCTGCCGACCGTCGACGTCTTTT
       P  I  R  P  V  R  T  A  R  P  H  R  L  P  T  V  D  V  F  F
3721 TTCCGGATTTTTACGGCGATGGACTGACGAAAGACCGTTTTATCTCGCTGATCGACCGAA
       P  D  F  Y  G  D  G  L  T  K  D  R  F  I  S  L  I  D  R  I
3781 TTGGCGACGGGCAGACGGCGGAAGTGATGTGCCACCCGGCGTATATCGATGTTCCGCTCG
       G  D  G  Q  T  A  E  V  M  C  H  P  A  Y  I  D  V  P  L  A
3841 CGTCAGGAAGTTCCTATTGTCAACAACGAGTCGAAGAGCTGGCGGTGCTGACCGACCCAA
       S  G  S  S  Y  C  Q  Q  R  V  E  E  L  A  V  L  T  D  P  T
3901 CGCTCGTTGCCGAGATGGCCGAGCGCGGTGTTCAGCTGATCACGTACCGCGAATTCTATA
       L  V  A  E  M  A  E  R  G  V  Q  L  I  T  Y  R  E  F  Y  K
           UCCUCcaCU
3961 AACTATAGGAGAGGGCGTTATGCAAACGTATGAACAAACTGTATTCCAACTGATTCTTCA
       L  *   RBS       M  Q  T  Y  E  Q  T  V  F  Q  L  I  L  H
                        CelD →
4021 TGGCGGAAACGGCCGCAGTTATGCAATGGAGGCGATTACGGCGGCGAAAAAAGGGGAATT
      G  G  N  G  R  S  Y  A  M  E  A  I  T  A  A  K  K  G  E  F
4081 TGCCGAGGCGCGCAGGCTGCTTGAACAGGCGGGAGCAGAACTGCAGGCGGCCCATGGGCT
      A  E  A  R  R  L  L  E  Q  A  G  A  E  L  Q  A  A  H  G  L
4141 GCAGACCGCGTTGCTGCAACAAGAAGCGAGCGGCGGGCAGCCGGTGGTGACGCTTCTGAT
      Q  T  A  L  L  Q  Q  E  A  S  G  G  Q  P  V  V  T  L  L  M
4201 GGTGCATGCCCAAGATCATTTAATGACGGCGATCACGGTCAAGGATTTAGCCGCTGAATT
      V  H  A  Q  D  H  L  M  T  A  I  T  V  K  D  L  A  A  E  F
4261 CGTGGAGCTGTATGAAGCGCTCAAGCGGCAAACAACCGAATCATAATTGTCGTCGGCCGG
      V  E  L  Y  E  A  L  K  R  Q  T  T  E  S  *
4321 CTTGTTGACCGGGTGAGGATGATTCGCTTCAACGGGAATGAACCCAAGGCGGGCGGCAAC
                                                     --------   -
4381 GGGCTGTTCTAGCGACCGCCCGCTTTGTTGATTTTACAATAATTTTATTATGTGAACATA
       →     ←  -   --------
4441 AGGGCGCCAGCCGTTCAGCCCTCCCGGCTGCCGCCGGTCTCGGCTGAACGGAGAAGGAAA
4501 ACTCCGCTTGTTTGCGTCACATATCGTGGCTTGTGTTGTGCTTTTGCCGCGTGTGATTGC
```

```
4561 GGTTCTTCACTTTTTTCGAACCGTCAAGCGGCTCCGGCTGCCCAGGATTGTCGCCTTTTG
4621 GGGCGTTTTTGCGCATATCTTTGCCGTCGTTTTTGTTCGTCATTTGTTGTCCCTCCTTTG
4681 TCCGTTAGCATGCGGCGTTGTGCTTTTGCTTATTCGGGGCATAATTTGCGCCGGTTTGAA
4741 ACGATAAGTAGTGGATGACCGCGATGACTGCGAAGGAGGAACGAGCCATGGTTTATCATA
4801 AAACGAAGCAAGACGCGTTTCAGGCGCGCGCAAAAGGCGACGATGGAGGCAAAAGAGTGG
4861 CATGACCATCTCGTCCGCGATCAAGCCGATTACGGCCATCAGTTAGCCCATTTGCGGCAG
4921 GAAGTGAACGAAGCGTTTGCACAAATTGAAAACGCGCTTGAAGTCGCCTCGGAAACGCAG
4981 CGCATGCAGCTTGAGAAATTCCGCAGCGATTTGCAGGCGATCCCCGG
```

possible *rho*-independent terminator at the 3' end (Rosenberg and Court, 1979). The coding regions for all four genes utilize ATG codons to initiate transcription. The potential ribosomal-binding sequences for *celA* and *celC* provide a better match with the 3' terminus of 16s rRNA from *B. stearothermophilus* (Van Charldorp et al., 1981). It is interesting to note that the potential ribosomal binding sequences in *celB, celC*, and *celD* are overlapped by stop codons from the preceding genes, and may provide translational coupling. A similar phenomenon was known in *B. subtilis* sucrose-specific PTS (Fouet et al., 1987), in which the stop codon for enzyme II overlapped the start codon for sucrose-6-phosphate-hydrolase.

Identification of Cloned Genes by Homology

An initial search of the DNA data base revealed strong homology between the translated sequences of the cloned genes and previously described genes encoding sugar-specific proteins in PTS systems (Table 8). The highest degree of similarity was observed with the *E. coli cel* operon (Table 8 and Fig. 5). Since MUC and MUG can be considered analogues of cellobiose, the sequenced genes from *B. stearothermophilus* XL-65-6 were designated as a *cel* operon, the first PTS cellobiose system to be characterized at a molecular level in a Gram positive organism. Table 9 summarizes some of the properties of these four genes. The gene products from *celA* and *celB*

Table 8. Comparison of predicted amino acid sequences for *celA, B*, and *D* and *celR'* from *B. stearothermophilus* to homologous PTS polypeptides from other organisms.

	%Identity (%Similarity)			
Proteins	*Bs*-CelA[a]	*Bs*-CelB[a]	*Bs*-CelD[a]	*Bs*-'CelR[a]
EIIBC				
Ec-CelA[b]	42.9 (59.2)			
Ll-LacE[c]	23.2 (50.5)[j]	30.6 (58.2)[k]		
Sa-LacE[d]	21.9 (51.0)[j]	29.4 (57.7)[k]		
Lc-LacE[e]	16.0 (48.0)[j]	29.1 (55.9)[k]		
Ec-CelB[b]		32.9 (60.1)		
EIIA				
Ec-CelC[b]			41.6 (67.3)	
Ll-LacF[c]			41.8 (66.0)	
Sa-LacF[d]			38.4 (62.6)	
Lc-FIII[e]			36.5 (58.9)	
Regulatory proteins				
Bsu-LevR[f]				29.5 (50.1)
Bsu-SacY[g]				27.2 (57.0)
Bsu-SacT[h]				30.0 (58.0)
Ech-ArbG[i]				30.2 (54.2)

[a] *Bs*-CelA, *Bs*-CelB, and *Bs*-CelD refers to polypeptides for gene *celA, celB*, and *celD* of *cel* operon from *B. stearothermophilus*. *Bs*-'CelR refers to the incomplete regulatory protein.

[b] *Ec*-CelA, *Ec*-CelB, and *Ec*-CelC refer to EIIB^{Cel}, EIIC^{Cel}, and EIIA^{Cel} of glucoside-specific PTS enzyme II of *cel* operon from *E. coli* (Parker and Hall, 1990; Reizer et al., 1990).

[c] *Ll*-LacE and *Ll*-LacF refer to lactose-specific PTS operon EIIBC^{Lac} and EIIA^{Lac} from *Lactococcus lactis*, respectively (de Vos et al., 1990).

[d] *Sa*-LacE and *Sa*-LacF refer to lactose-specific PTS EIIBC^{Lac} and EIIA^{Lac} of *lac*-operon from *Staphyloccus aureus*, respectively (Breidt et al., 1987).

[e] *Lc*-LacE and *Lc*-FIII refer to lactose-specific PTS EIIBC^{Lac} and EIIA^{Lac} from *Lactobacillus casei* respectively, (Alfred and Chassy, 1988; Alfred and Chassy, 1990).

[f] *Bsu*-LevR refers to transcriptional regulator of PTS levanase operon of *B. subtilis* (Débarbouille et al., 1991). The result of comparison with *Bs*-'celR were from the comparison of amino acid 2 to 459 with amino acid 475 to 935 of *levR* gene product.

[g] *Bsu*-SacY refers to antiterminator of PTS levansucrase gene of *B. subtilis* (Zukowski et al., 1990). The result of comparison with *Bs*-'celR were from the comparison of amino

Table 8 -- continued

acid 2 to 115 with amino acid 68 to 181 of *sacY* gene product.

[h] *Bsu*-SacT refers to transcriptional antiterminator of PTS sucrase operon of *B. subtilis* (Debarbouille et al., 1990). The result of comparison with *Bs*-'CelR were from the comparison of amino acid 2 to 101 with amino acid 66 to 165 of *sacT* gene product.

[i] *Ech*-ArbG refers to antiterminator of PTS phospho-ß-glucosidase operon of *Erwinia chrysanthemi* (El Hassouni, 1992). The result of comparison with *Bs*-'celR were from the comparison of amino acid 2 to 97 with amino acid 69 to 164 of *arbG* gene product.

[j] These results are comparisons of *Bs*-CelA with carboxyl terminal portions of *Ll*-LacE, *Sa*-LacE, and *Lc*-LacE, respectively.

[k] These results are comparisons of *Bs*-CelB with amino terminal portions of *Ll*-LacE, *Sa*-LacE, and *Lc*-LacE, respectively.

Figure 5. Comparison of deduced amino acid sequences of *cel* operon genes from *B. stearothermophilus* and *E. coli*. Identical residues are indicated by asterisks. Conserved residues with similar properties are indicated by dots. The proposed phosphorylation sites are overlined. Gaps (dashes) have been introduced to optimize alignment.

A. EIIB

```
Bs-CelA   M---NILLIC̄AAGMSTSLLVTKMKEAAKQKGIEANIWAVSADEAKSHLDQADVVLIGPQIRYKLAAFKKE  67
Ec-CelA   MEKKHIYLFCSAGMSTSLLVSKMRAQAEKYEVPVIIEAFPETLAGEKGQNADVVLLGPQIAYMLPEIQRL  70
          *   .* *.*.*********.**.  * .  .   * * .   *    ..*****.**** * *. ...
Bs-CelA   GEARGIPVDVINPADYGRVNGAGVLDFALR-LKK----                                  100
Ec-CelA   LPNK--PVEVIDSLLYGKVDGLGVLKAAVAAIKKAAAN                                  106
             .  **.**..  **.*.* ***  *.  .**
```

B. EIIC

```
Bs-CelB   MDRFIRMLEDRVMPVAGKIAEQRHLQAIRDGIILSMPLLIIGSLFLIVG--FLPIP--GYNEWMAKWFGE  66
Ec-CelB   MSNVIASLEKVLLPFAVKIGKQPHVNAIKNGFIRLMPLTLAGAMFVLINNVFLSFGEGSFFYSLGIRLDA  70
          *   *  **  ..* * **. * *..**..*.*  *** . *..*...   **..   ..   .. ...
Bs-CelB   HWLDKL--LYPVGA-----TFDIMALVVSFGVAYRLAEKYKVDALSAGAISLAAFLLATPYQVPFTPEGA  129
Ec-CelB   STIETLNGLKGIGGNVYNGTLGIMSLMAPFFIGMALAEERKVDALAAGLLSVAAFMTVTPYSVG------  134
            .. *  *  .*.     *..**.*. .* ..  ***  *****.** .*.***.  *** *
Bs-CelB   KETIMVSGGIPVQWVGSKGLFVAMILAIVSTEIYRKIIQKNIVIKLPDGVPPAVARSFVALIPGAAVLVV  199
Ec-CelB   -EAYAVGA----NWLGGANIISGIIIGLVVAEMFTFIVRRNWVIKLPDSVPASVSRSFSAF--NSRLYYS  197
           *.  *..    .*.*.  .. ..*...* .*..  *...* ******.**..*.*** *.   . .
Bs-CelB   VWVARLILEMTPFES-FHNIVSVLLNKPLSVLGGSVFGAIVAVLLVQLLWSTGLHGAAIVGGVMGPIWLS  268
Ec-CelB   FRDGDYCLALNTWGTNFHQIIMDTISTPLASLGSVWLAYVI--LSTALVLRIHAACADRTGQRHYDAW-A  264
           . .   * .    . **.*.   .. **. **.  .. ..  *   *.       *   *      * .
Bs-CelB   LMDENRMVFQQNPNAELPNVITQQFFDLW--------IYIGGSGATLALALTMMFRARSRQLKSLGRLAI  330
Ec-CelB   L--ENIATYQQYGSVE-AALAAGKTFHIWAKPMLDSFIFLGGSGATLGLILAIFIASRRADYRQVAKLAL  331
          *  **   .**  . * . . . . *..*        *..*******.* *.. . .*  . . ...**.
Bs-CelB   APGIFNINEPITFGMPIVMNPLLIIPFILVPVVLVVVSYAAMATGLVAKPSGVAVPWTTPIVISGYLATG  400
Ec-CelB   PSGIFQINEPILFGLPIIMNPVMFIPLYWYNRILAAITLAAYYMGIIPPVTNIA-PWTMPTGLGAFFNTN  400
          ..***.***** **.**.***...**.     .*  .. **   * ... . .* *** *  ..... *
Bs-CelB   GKISGSILQIVNFFIAFAIYYPFFSIWDKQKAAEEQADPTISSGAGATHSL                     451
Ec-CelB   GTSPHC--------------------WSHSSTLASQR--------------                     417
          *  .                      *    .   *
```

C. EIIA

```
Bs-CelD   M-------------QTYEQTVFQLILHGGNGRSYAMEAITAAKKGEFAEARRLLEQAGAELQAAHGLQTA  57
Ec-CelC   MMDLDNIPDTQTEAEELEEVVMGLIINSGQARSLAYAALKQANRGDFAAAKAMMDQSRMALNEAHLVQTK  70
          *             .  *. *  **...*..** *  *.  *..*.** *. ...*.   *. ** .**
Bs-CelD   LLQQEASGGQPVVTLLMVHAQDHLMTAITVKDLAAEFVELYEALKRQTTES                     108
Ec-CelC   LIEGDAGEGKMKVSLVLVHAQDHLMTSMLARELITELIELHEKLK-----A                     106
          *.. .*. *.  *.*..*********..  ..* .*..** * **     .
```

Table 9. Characteristics of the *cel* operom from *B. stearothermophilus* XL-65-6

Genes	PTS Component	No. of Nucleotides	%GC	Total No. of Amino Acids	M_r	Charges	Predicted pI
celA	EIIB^{Cel}	303	53	100	10,737	+3	9.8
celB	EIIC^{Cel}	1,356	53	451	48,805	+4	10.0
celC	Putative cleavage enzyme	738	59	245	27,423	-8	6.1
celD	EIIA^{Cel}	327	55	108	11,785	-5	5.2

have high pI values and positive charges while those from *celC* and celD have low pI values and negative charges.

The translated amino acid sequence for *B. stearothermophilus celA* shares 43% identity and 59% similarity with that of *celA* from *E. coli* K12 (Parker and Hall, 1990) which encodes a part of enzyme II (EIIB) (Reizer et al., 1990) (Table 8 and Fig. 5A). In many organisms, the corresponding domains encoded by *celA* and *celB* are fused into a single polypeptide (EIIBC) (Saier and Reizer, 1992). The translated *B. stearothermophilus celA* sequence is homologous to the carboxyl terminal portion (EIIB domain) of the lactose-specific enzyme II proteins from *Lactococcus lactis, Staphylococcus aureus,* and *Lactobacillus casei* (Table 8). Cys^7 is surrounded by a highly conserved region in all homologues examined. This Cys residue has been proposed as the site of phosphorylation in *L. lactis* (Alfred and Chassy, 1990). The *celA* gene is predicted to encode a protein with M_r 10,737 (Table 9).

The *celB* gene begins 17 bases downstream from *cel*A and is preceded by a potential ribosomal-binding site which overlaps the last 2 bases of the *celA* stop codon. The *celB* gene is predicted to encode a protein with M_r 48,805. The deduced amino acid sequence for *celB* is most similar to that of the membrane-spanning polypeptide in the enzyme II complex (EIIC) for cellobiose from *E. coli* (*celB*) (Parker and Hall, 1990), exhibiting 33% identity and 60% similarity (Table 8 and Fig.

5B). The translated sequence for *B. stearothermophilus cel B* is also similar to the amino terminal portion of genes encoding PTS enzyme II proteins from other organisms (Table 8).

The *celC* gene begins 14 bases downstream from the *celB* stop codon and also exhibits an overlap between the ribosomal-binding region and the preceding stop codon. In contrast to the other genes in the *cel* operon, no genes were found in the current data base with similarity to the translated sequence for *celC*. However, the carboxyl-terminal portion of this protein exhibited partial homology (22% identity, 53% similarity) with the carboxyl portion of *E. coli celF* encoding the phospho-ß-glucosidase. Thus the *celC* gene is presumed to encode a cleavage enzyme for cellobiose-phosphate.

The *celD* gene encodes a small protein containing 108 amino acids. Again, the proposed ribosomal binding site overlaps the stop codon of the preceding gene. The translated amino acid sequence for *celD* exhibits 41.6% identity and 67.3% similarity to that of *E. coli celC* which encodes enzyme III (EIIA) (Table 8 and Fig. 5C). Table 8 summarizes the homology between *celD* and related genes from other organisms. A histidine residue (His^{76}) is located in a region which is highly conserved among different organisms. This residue has been proposed as the site of phosphorylation for the lactose-specific enzyme III in *L. casei* (Alfred and Chassy, 1988; Hengstenberg et al., 1989).

An incomplete open reading frame (Fig. 3 and 4) was found upstream from the *cel* operon which contained 463 codons and is transcribed in the same direction as the *cel* operon. This gene was provisionally identified as encoding a regulatory protein (*celR*) based on homology to regulatory proteins in other PTS systems (Table 8). The translated sequence for the incomplete *celR* gene is most similar to that of *B. subtilis levR* (de Vos et al., 1990), the transcriptional regulator for the PTS levanase operon. Although a putative promoter for *celABCD* was readily evident in the intercistronic sequence, it is unclear whether *celR* is part of the *cel* operon or transcribed separately. No stem-loop structure resembling a *rho*-independent terminator was found in the intergenic region and it is possible that all five genes may be transcribed at the some level from a common promoter.

A portion of the amino terminus of a possible sixth gene was found downstream from the *cel* operon but was not identified by homology searches of the current DNA data base.

Codon Usage

Codon usage in the *cel* genes was compared to that for the genes encoding phosphoglycerate kinase (Davies et al., 1991) and alcohol dehydrogenase T (Sakoda and Imanaka, 1992) from *B. stearothermophilus* (Table 10). Many codons (ie. AGA, TGA, TAG, ATA, and TGT) were rarely used. Either a G or C was usually present in the wobble positions resulting in G+C contents of 52.8 to 54.9 for *celAB and celD*, respectively.

Table 10. Codon Usage of *cel* Operon

		Mole %					
AmAcid	Codon	*cel*A	*cel*B	*cel*C	*cel*D	*gap*	*adh*T
Gly	GGG	2.00	2.19	2.57	2.18	0.62	1.82
	GGA	1.04	2.02	0.88	1.85	1.17	1.82
	GGT	1.04	0.93	0.88	0.00	0.28	4.70
	GGC	4.00	3.29	3.01	2.81	4.81	2.35
Glu	GAG	4.00	0.67	4.33	2.75	1.17	0.61
	GAA	1.00	2.66	3.01	7.44	5.70	6.21
Asp	GAT	1.98	1.34	3.03	1.85	2.09	3.58
	GAC	4.02	1.10	2.69	0.00	3.88	1.46
Val	GTG	4.00	3.57	2.87	3.72	4.18	2.35
	GTA	0.00	0.71	0.45	0.94	0.36	5.09
	GTT	2.00	2.24	2.42	0.00	2.03	1.44
	GTC	2.00	3.77	3.23	0.94	5.37	4.18
Ala	GCG	6.00	4.69	4.11	9.29	5.70	1.82
	GCA	1.05	0.42	0.34	1.89	1.16	3.53
	GCT	4.05	0.94	0.77	0.94	1.16	2.93
	GCC	4.05	4.48	3.34	3.78	3.61	1.82
Arg	AGG	0.00	0.00	0.41	0.93	0.00	0.00
	AGA	0.00	0.00	0.00	0.00	0.00	0.00
	CGG	2.00	1.77	3.67	0.93	0.31	0.00
	CGA	0.00	0.22	1.22	0.00	0.31	0.00
	CGT	1.00	0.22	0.82	0.00	0.31	1.18
	CGC	1.00	0.90	2.04	1.85	3.58	1.50
Ser	AGT	2.00	0.90	0.37	0.92	0.00	0.00
	AGC	0.00	1.80	0.37	0.92	0.59	0.29
	TCG	1.00	1.14	1.84	0.00	3.87	0.59
	TCA	1.00	0.90	0.73	0.92	0.00	0.88
	TCT	0.00	0.42	0.00	0.00	0.32	0.29
	TCC	0.00	0.90	0.37	0.00	0.59	0.59
Lys	AAG	4.00	1.10	0.54	1.85	0.89	1.16
	AAA	6.00	1.38	1.09	1.85	5.97	7.14
Asn	AAT	1.00	1.33	1.23	0.00	0.60	0.59
	AAC	3.00	1.33	1.63	0.93	5.37	1.78
Met	ATG	3.00	3.77	2.04	3.70	2.39	0.89

Table 10 -- continued

		Mole %					
AmAcid	Codon	*cel*A	*cel*B	*cel*C	*cel*D	*gap*	*adh*T
Ile	ATA	0.00	0.00	0.36	0.00	0.00	0.54
	ATT	5.04	6.34	1.89	1.86	2.09	3.56
	ATC	3.04	3.77	2.25	0.92	3.88	1.78
Thr	ACG	2.00	2.00	3.28	4.67	3.57	0.29
	ACA	0.00	0.68	0.00	0.92	0.62	2.68
	ACT	0.00	0.00	0.83	0.92	0.00	0.29
	ACC	0.00	1.09	0.83	1.83	1.47	0.00
Trp	TGG	1.00	2.22	0.00	0.00	0.60	1.19
End	TGA	0.00	0.00	0.00	0.00	0.00	0.00
	TAG	0.00	0.00	0.00	0.00	0.00	0.00
	TAA	1.00	1.00	1.00	1.00	1.00	1.00
Cys	TGT	0.00	0.00	0.67	0.00	0.00	0.59
	TGC	1.00	0.00	1.37	0.00	0.60	2.08
Tyr	TAT	2.00	1.34	2.09	2.78	0.60	2.09
	TAC	0.00	1.10	1.58	0.00	1.79	1.18
Leu	TTG	3.00	3.51	1.56	0.91	5.64	0.31
	TTA	1.00	0.90	0.00	1.81	0.56	3.86
	CTG	2.00	3.51	4.67	6.48	0.89	0.31
	CTA	0.00	0.23	0.45	0.00	0.00	0.00
	CTT	1.00	1.58	0.90	2.72	0.32	2.39
	CTC	3.00	1.58	1.71	0.91	0.56	0.93
Phe	TTT	1.00	2.88	3.98	0.92	0.60	1.18
	TTC	1.00	2.66	1.33	1.86	0.80	0.89
Gln	CAG	0.99	1.55	0.94	3.67	0.00	0.00
	CAA	2.01	1.55	1.91	6.52	0.90	3.56
His	CAT	1.00	0.67	2.60	3.70	2.10	2.67
	CAC	0.00	0.44	1.89	0.00	0.59	0.00
Pro	CCG	2.01	3.79	5.76	0.93	3.28	1.46
	CCA	0.99	1.35	1.18	0.00	0.00	0.91
	CCT	0.00	0.64	0.00	0.00	0.00	2.67
	CCC	0.00	0.64	0.00	0.00	0.00	0.00

The G+C content of *celC* is unusually high (59.1%) and differs significantly from other genes in this operon.

Southern hybridization analysis

Southern analysis confirmed that the *cel* genes originated from *B. stearothermophilus* XL-65-6. A *Sty*I fragment from the internal region of the *cel* operon was used as a probe. This probe did not hybridize to DNA fragments from *E. coli* DH5α and bound single regions in digestions of XL-65-6 genomic DNA, consistent with the *cel* operon being present as a single chromosomal copy (Fig. 6).

Hydropathy analysis

Analyses of the deduced amino acid sequences by the method of Kyte and Doolittle (1982) revealed a hydrophilic or slightly hydrophobic character for the products of *celA*, *celC*, and *celD* (Fig. 7). Further analysis for membrane spanning domains by the method of Klein et al. (1985) predicted a single domain in the *celA* product which may serve as a membrane anchor. No membrane-spanning domains were predicted in the *celC* and *celD* products. Hydropathy profiles for the *celA* and *celD* encoded polypeptides exhibit a remarkable degree of conservation when compared to homologous PTS enzymes from other organisms despite differences in the transported sugars (Fig. 7). The *celB* gene product is very hydrophobic with 9 to 11 predicted membrane-spanning domains. The hydropathy plot for this protein is also very similar to the *celB* homologues.

Reizer et al. (1990) previously noted that the N-terminal

Figure 6. Southern hybridization analysis of *cel* operon. Lane assignment: Lane 1 to lane 5: *sty*I, *pst*I, *Hind*III, *Eco*RI, and *Bss*HII digestions of *B. stearothermophilus* XL-65-6 genomic DNA; Lane 6: *pts*I digestion of *E. coli* DH5α genomic DNA; Lane 7: λ marker.

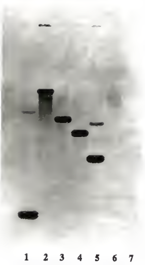
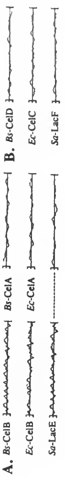

1
2
3
4
5
6
7

Figure 7. Hydropathy plots of PTS enzymes from different organisms. *Bs*-CelB and *Ec*-CelB, *Bs*-CelA and *Ec*-CelA refer to EIIC and EIIB from *B. stearothermophilus* (*Bs*) and *E. coli* (*Ec*), respectively. *Sa*-LacE refers to EIIBC from *Staphylococcus aureus*. *Bs*-CelD, *Ec*-CelC, and *Sa*-LacF denote EIIA from the 3 listed organisms.

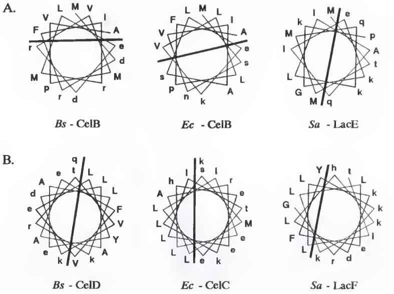

A.
Bs-CelB
Bs-CelA
Ec-CelB
Ec-CelA
Sa-LacE
B.
Bs-CelD
Ec-CelC
Sa-LacF

region of *E. coli* *celB* gene product and the C-terminal region of the *E. coli* *celC* translation product can be folded into helical wheel plots to reveal distinctly hydrophobic and hydrophilic faces. A similar analysis of the corresponding regions in the translated sequences for *B. stearothermophilus* *celB* and *celD* confirmed that these patterns were also maintained in this Gram positive organism (Fig. 8). When folded, the first 16 amino acids at the amino terminus of the *B. stearothermophilus* *celB* encoded protein produced a highly hydrophobic face which may be involved in the formation of the transport complex but is not predicted to span the membrane (Klein et al., 1985). Similarly, the carboxyl terminal region of the *celD* gene product (close to the histidinyl residue of the proposed phosphorylation site) can also be folded into an α-helix in which most hydrophobic residues are segregated on one side and may function in the assembly of the active enzyme II complex.

Functional analysis of genes in the *B. stearothermophilus* *cel* operon in *E. coli*

The *B. stearothermophilus* *cel* genes were constitutively expressed in strain DH5α(pLOI902). Recombinant cells hydrolysed MUC, MUG, and *p*-nitrophenyl-glucopyranoside (pNPG), and formed slight pink colonies when growing on cellobiose-MacConkey agar, indicating poor utilization of cellobiose. All activities were concurrently lost as a result of deletions in the carboxyl terminal region of *celD* and frame-shift mutations in *celB* (Fig. 3). The full insert was subcloned in

Figure 8. Comparisons of helical wheel plots of terminal regions from PTS enzymes. Hydrophobic residues are indicated by capital letters and hydrophilic residues by small letters. Hydrophobic and hydrophilic phases are divided by the line. *Bs*-CelB, *Ec*-CelB, and *Sa*-LacE are folded from the first 16 amino acid residues at amino terminal ends EIIC. *Bs*-CelD, *Ec*-CelC, and *Sa*-LacF are folded from the carboxyl terminal ends of EIIA, from amino acid 86 to 104, 97 to 115, and 86 to 103, respectively.


A.
L M V I A e d M r d r p M r F V
L M L I A e s L A k n p s V V F
I M e q p A t k k q M G L I M k
B.
q t L L L F V Y A k V k e A r e d A e
k s I r e t M e k e L L L L L A h I
Y h t L k k k I e d r k L F L G L L

Bs - CelB *Ec* - CelB *Sa* - LacE

Bs - CelD *Ec* - CelC *Sa* - LacF

the reverse orientation to provide transcription from the *lac* promoter. Deletions in the amino terminal region of *celA* also resulted in a loss of activity (Fig. 3). A variety of approaches were tried to isolate mutations in the *celC* gene without success. Thus a requirement for three of the four genes within the *cel* operon has been demonstrated while the fourth gene, *celC*, is presumed to encode a cleavage enzyme for cellobiose-phosphate.

Cells of strain DH5α(pLOI902) exhibited strong MUC, MUG and pNPG activities when incubated at 37°C. However, little activity was detected when cells were incubated at 55°C despite the thermal tolerance of *B. stearothermophilus*. Previous studies had demonstrated that the general proteins (EI and HPr) of the *E. coli* PTS complement sugar-specific components from other Gram positive bacteria (de Vos et al., 1990; Fouet et al., 1987). The loss of activity at 55°C could result from thermal inactivation of essential host-supplied components (enzyme I). To test this hypothesis, a *ptsI* mutant of *E. coli*, MM6 (Fraenkel et al., 1964), was transformed with pLOI902. The resulting recombinant was inactive at both temperatures.

The requirement for host EI (and presumably also HPr) for activity at both temperatures indicates that the phosphorylation of chromogenic substrates may be essential prior to cleavage.

Discussion

The chromogenic substrate MUC has often been used to measure exocellobiohydrolase activity, although potential confounding activities from ß-glucosidases are well known (Coughlan, 1988). However, the discovery that MUC and MUG activities could be used as a marker for the cloning of a cellobiose PTS operon from a Gram positive bacterium, *B. stearothermophilus*, was rather unexpected.

The organization of the PTS has been described by many researchers (Saier and Reizer, 1992). This system consists of two general energy-coupling proteins, enzyme I and HPr, as well as a sugar-specific permease, commonly referred as the enzyme II complex. Although the enzyme II complex may consist of 1, 2, 3, or 4 distinct polypeptide chains, each complex contains at least three functional domains: a hydrophobic trans-membrane domain which binds and transports the sugar, a closely associated hydrophilic domain which possesses the first phosphorylation site, and a second hydrophilic domain containing an additional phosphorylation site. In *B. stearothermophilus*, *celB* encodes the membrane-spanning polypeptide which forms the transmembrane channel. The *celA* product is less hydrophobic and contains a hydrophobic tail which may serve as a membrane anchor. The *celA* product is predicted to have a high pI, similar to that of the *celB* product. Since *celA* and *celB* encode domains which form a single polypeptide in many organisms (Saier and Reizer, 1992), these two gene products can be assumed to interact closely in the enzyme II complex. The *celD* product is hydrophilic although a hydrophobic surface appears to be present near the carboxyl terminus. This gene product is predicted to have a low pI which may allow the formation of ionic interactions with both the *celA* and *celB* encoded polypeptides. The *celC* product is also very hydrophilic and is proposed to encode an enzyme for the cleavage of cellobiose-phosphate, based on homology with *E. coli celF*. All attempts to mutate this gene were unsuccessful.

In recombinant *E. coli* expressing MUC and MUG activities, enzyme I (and presumably HPr) must be supplied by the host and complement the PTS genes from *B. stearothermophilus*. Other examples of functional complementation with heterologous PTS systems from Gram positive organisms have been reported previously in recombinant *E. coli* (de Vos et al., 1990; Fouet et al., 1987).

MUC, MUG, and pNPG are analogues of cellobiose and require hydrolysis to release the respective chromogens. Since a functional *E. coli* enzyme I (*ptsI*) was required for hydrolysis, phosphorylation can be presumed to be essential. In the simplest case, the phosphorylated intracellular product must be hydrolysed once to release the chromogen from MUG or twice to release the chromogen from MUC. However, it is possible that the transported substrate is cleaved by a phosphorylase as proposed for *Cellulomonas uda* and *C. favigena*

(Coughlan and Mayer, 1992). A phosphorylase would produce 4-methylumbelliferyl-phosphate or p-nitrophenyl-phosphate as products which must be cleaved by a host phosphatase to release the chromogen. The precise nature of the reaction catalyzed by the proposed cleavage enzyme (*celC*) is still unknown.

The PTS cellobiose-specific proteins from the Gram positive thermophile *B. stearothermophilus* (*celA*, *celB*, and *celD*) appear remarkably similar to the corresponding genes from the Gram negative mesophile, *E. coli celA, celB*, and *celC* (Parker and Hall, 1990) and may be derived from a common ancestor. Indeed, the sugar-specific proteins from most organisms which have been examined exhibit a high degree of homology despite differences in the transported substrates (Saier and Reizer, 1992). However, the proposed phospho-ß-glucosidases from these two organisms (*celC* in *B. stearothermophilus* and *celD* in *E. coli*, respectively) exhibit only modest homology to each other and little homology to other glycohydrolases, consistent with an independent origin. The independent origin of *celC* is further supported by the high G+C composition of this gene, considerably higher than that of other sequenced genes from *B. stearothermophilus*. The discovery of a PTS *cel* operon in *B. stearothermophilus* which is similar to that in *E. coli* (Parker and Hall, 1990) is consistent with the PTS being the prevalent route for cellobiose uptake in bacteria.

Transcription of *B. stearothermophilus cel* operon may be regulated by a mechanism comparable to that for *lev* operon in *B. subtlis*. Transcription of *B. subtilis lev* operon, a PTS levanase operon, is controlled by a regulator, LevR. LevR functions as a repressor in the uninduced state and as an activator after induction with fructose. The gene encoding this regulator is located upstream of the native promoter of *lev* operon (Débarbouillé et al., 1991). *B. stearothermophilus* 'CelR exhibits homology to carboxyl terminal portion of *B. subtilis* LevR, and genes encoding these two enzymes are arranged in a comparable pattern. Therefore, CelR is expected to have a similar function to LevR. 'CelR is also homologous to other transcriptional antiterminators from *B. subtilis* (Debarbouille et al., 1990; Zukowski et al., 1990). Thus CelR may function as a transcriptional antiterminator in *B. stearothermophilus*.

CHAPTER 5
CLONING AND SEQUENCING OF THE PHOSPHOENOLPYRUVATE-DEPENDENT PHOSPHOTRANSFERASE GENES (*PTSHIT*) FROM *BACILLUS STEAROTHERMOPHILUS* AND HYPEREXPRESSION IN *ESCHERICHIA COLI*

The phosphoenolpyruvate-dependent phosphotransferase system (PTS) is a complex carbohydrate uptake system which has a variety of physiological functions in bacteria (Postma et al., 1993; Saier and Peizer, 1992). This system consists of two cytoplasmic energy-coupling proteins, enzyme I (EI) and histidine-containing phosphocarrier protein (HPr), and sugar-specific permeases referred to as enzyme II complexes (EII). The PTS functions to transfer a phosphoryl group from phosphoenolpyruvate (PEP) to EI, to HPr, and then to the EII complex. The EII complex catalyzes carbohydrate uptake with concomitant phosphorylation.

In the previous chapter, a cellobiose PTS operon (*cel* operon) from an isolate of *Bacillus stearothermophilus*, strain XL-65-6 was described. Cellobiose utilization by *Escherichia coli* recombinants harboring this operon was poor (slight acidification of MacConkey agar containing cellobiose) although clones were easily identified using chromogenic cellobiose analogues. Since EIs from different organisms have been reported to complement poorly in heterologous systems (Postma et al., 1993), the inability of recombinant *E. coli* to utilize cellobiose efficiently may result from poor compatibility between the native general proteins (EI and HPr) and the cellobiose-specific *B. stearothermophilus* gene products.

The aim of this objective was the cloning and sequencing *ptsH* and *ptsI* genes from *B. stearothermophilus* strain XL-65-6, and improvement of cellobiose utilization in recombinant *E. coli* by combining *cel* operon and *ptsHI* operon.

Materials and Methods

Bacterial Strains, Plasmids, and Media

Bacterial strains and plasmids used in this study are listed in Table 11. Luria broth (or agar) were used for growth and maintenance of *E. coli* (Atlas and Parks, 1993). Acidification of Difco MacConkey agar base plus 1.0% carbohydrate was used as a test for carbohydrate utilization. Complementation of *E. coli ptsI* mutations was evaluated using MacConkey agar containing each of 5 PTS carbohydrates (fructose, glucose, mannose, mannitol, or sorbitol). Sorbitol-MacConkey agar was used to test functional complementation of the *ptsH* mutation in *E. coli* 1101. Carbohydrate utilization was also examined in M9 minimal medium supplemented with 0.2% carbohydrate (Atlas and Parks, 1993). Ampicillin (50 $\mu g\ ml^{-1}$) was added to the media as appropriate for the selection of recombinants.

Table 11. *E. coli* strains and plasmids used in this study

Strains or plasmids	Genetic characteristics	Source/reference
E. coli		
DH5α	F$^-$ Δ*lacZM15 deoR recA1 endA1 supE44* *hsdR17*(r_{K-}, m_{K}^{+}) λ$^-$ *thi-1 gyrA96 relA1*	BRL[a]
MM6	*lacI22 dctB3 ptsI2 relA1 thi-1 spoT1*	Fraenkel et al., 1964
1101	*his-64 ptsH1 relA1 spoT1 thi-1 bglR11*	Fox and Wilson, 1968
JLT2	F$^-$ *mcrB mrr hsdS20*(r_{B}^{-}, m_{B}^{-}) *recA13* *leu mtl1 supE44 ara14 galK2 lacY1* *proA2 rpsL20*(Sm^r) *xyl5* λ$^-$ *ptsI*	Tigemeyer, 1986
BL21	Δ*ptsI*	Chauvin[b]
Plasmids		
pUC18	*bla lacI'Z'*	BRL
pLOI1801	pUC18 containing *pts'G ptsHIT wapA*	This study
pLOI1803	pUC18 containing *ptsHI*	This study
pLOI1804	pUC18 containing *ptsHI*	This study
pLOI1805	pUC18 containing *pts'G ptsH*	This study
pLOI1806	pUC18 containing *ptsHIT wapA'*	This study
pLOI1807	pLOI1806 with mutated *ptsH*	This study
pLOI1808	pUC18 containing *ptsHI* and *cel* operon	This study
pLOI902	pUC18 containing *cel* operon	Previous study

[a] Bethesda Research Laboratories, Gaitherburg, Md.

[b] Department of Biology, Johns Hopkins University.

DNA Manipulation

Standard procedures were used for the construction, isolation, and analysis of plasmids (Sambrook et al., 1989). Commercial restriction enzymes were used as recommended by each manufacturer.

Cloning of *B. stearothermophilus ptsHIT* Operon

Clones containing the XL-65-6 *ptsH* and *ptsI* genes were isolated by transforming an *E. coli ptsI* mutant, strain MM6 (Fraenkel et al., 1964), with a previously constructed pUC18-based genomic library. Transformants were screened for the utilization of PTS sugar by plating on MacConkey agar containing fructose. Deep red colonies were selected after overnight incubation at 30°C and purified for further investigation. *E. coli* DH5α was used as the host for subcloning experiments.

Southern Hybridization

Restriction endonuclease digestions of genomic DNA from strain XL-65-6 were prepared using *Cla*I, *Eco*RI, *Pst*I, *Sal*I, and *Sma*I. After separation by agarose gel electrophoresis, fragments were transferred to Zeta-Probe GT membranes (Bio-Rad Laboratories, Richmond, Calif.) and probed with a 1.3-kbp *Eco*RI fragment spanning parts of *ptsH* and *ptsI* from *B. stearothermophilus* XL-65-6. This probe was labeled with digitoxigenin (Genius System; Boehringer Mannheim Biochemicals, Indianapolis, Ind.) and hybridized at 55°C

(Sambrook et al., 1989). Genomic DNA from *E. coli* DH5α was digested with *Pst*I and served as a negative control.

DNA Sequencing and Sequence Analysis

DNA sequencing and sequence analysis were as described in the previous chapter. The nucleotide sequence data has been submitted to GenBank and assigned the accession number U12340.

Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Cultures of *E. coli* JLT2 harboring recombinant plasmids were grown in Luria broth with shaking at 30°C and 37°C overnight. Cells were harvested by centrifugation and washed twice with 10 mM sodium phosphate buffer (pH 7.0). Cell pellets were resuspended in an equal volume of this buffer and disrupted by two passes through a French pressure cell at 20,000 lb in^{-2}. Cell debris was removed by centrifugation (13,000 X g, 10 min). Nucleic acid was digested by adding $CaCl_2$ (2 mM final concentration) and 50 μg ml^{-1} DNase (1 h, 22°C). Denatured proteins were separated by SDS-PAGE (Schägger and von Jagow, 1987) and stained with Coomassie blue.

Results

Cloning of the *ptsHIT* Operon from *B. stearothermophilus* XL-65-6

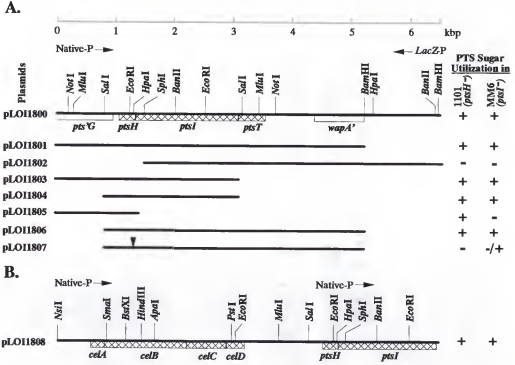
Clones containing *B. stearothermophilus* XL-65-6 *ptsI* gene were readily isolated by functional complementation of a *ptsI* mutation in *E. coli* MM6 provided that the MacConkey agar selection plates containing ampicillin were incubated at 30°C.

Initial attempts using MacConkey agar with glucose or fructose for the cloning were not successful when the culture was incubated at 37°C. Plasmids purified from all six positive clones were stable in *E. coli* DH5α and restored the ability of MM6 to utilize 5 PTS carbohydrates at 30°C. Although these MM6 recombinants also formed red colonies at 37°C indicating complementation, growth was severely limited. Based upon an analysis of restriction sites, plasmids in 5 clones appeared identical, with a DNA insert of 6.4 kbp, while the other was slightly smaller (5 kbp). One of the larger plasmids, denoted pLOI1800, was selected for further study (Fig. 9A).

Transformation of pLOI1800 into a *ptsH E. coli* mutant, strain 1101, also restored PTS function of this mutant indicating the presence of *ptsH* gene in pLOI1800.

A series of subclones was constructed to localize these genes (Fig. 9A). Plasmid pLOI1801 was constructed from pLOI1800 by deleting the *Bam*HI to *Sac*I (polylinker region) fragment and retained both activities. Both activities were lost after deletion of the *Pst*I (polylinker region) to *Sph*I fragment (pLOI1802). Plasmid pLOI1803, an exonuclease III deletion of pLOI1800, retained both activities. After deletion of the terminal *Sal*I (one *Sal*I site in the polylinker region) fragment in pLOI1801, both activities were retained (pLOI1806). By making a similar *Sal*I deletion in pLOI1803 to produce pLOI1804, the coding regions for both enzymes were localized to the 2.3 kbp region. EI activity was eliminated

Figure 9. Diagrams illustrating the features of recombinant plasmids. The ability of constructs to complement the *ptsH* mutation present in *E. coli* 1101 and the *ptsI* mutation present in *E. coli* MM6 are indicated on the right, a plus sign being affirmative. The -/+ indicates a weak positive reaction (light pink colonies) and a minus sign indicates lack of complementation. **A.** Restriction map of *B. stearothermophilus* DNA fragment in pLOI1800 and subclone. Plasmid pLOI1800 contains a *Mob*I fragment which has been cloned into the *Bam*HI site of pUC18. Polylinker sites on either side are not shown. The arrows indicate the direction (but not the location) of the *lacZ* (within vector) and native *B. stearothermophilus* promoters. The hatched area corresponds to the *ptsHIT* operon. Subclones are aligned beneath the corresponding sites in pLOI1800. The position of the frameshift mutation in pLOI1807 is indicated by an arrowhead. **B.** Plasmid pLOI1808 which includes *B. stearothermophilus* genes encoding the PTS general proteins (*ptsHI*) and the cellobiose specific PTS permease (*celABCD*).


A.
0
1
2
3
4
5
6
kbp
Native-P
LacZ-P
Plasmids
Not I
Mlu I
Sal I
EcoRI
HpaI
SphI
BanII
EcoRI
SalI
MluI
Not I
BamHI
HpaI
BanII
BamHI
pts'G
ptsH
ptsI
ptsT
wapA'
PTS Sugar Utilization in
1101 (ptsH⁻)
MM6 (ptsI⁻)
pLOI1800 + +
pLOI1801 + +
pLOI1802 - -
pLOI1803 + +
pLOI1804 + +
pLOI1805 + -
pLOI1806 + +
pLOI1807 - -/+
B.
Native-P
Native-P
NsiI
SmaI
BstXI
HindIII
ApaI
PstI
EcoRI
MluI
SalI
EcoRI
HpaI
SphI
BanII
EcoRI
celA
celB
celC
celD
ptsH
ptsI
pLOI1808 + +

by deleting 5 kbp from pLOI1800 with exonuclease III, although HPr function was retained (pLOI1805).

Using the internal 1.3 kbp *Eco*RI fragment (Fig. 9A) from pLOI1800 as a probe, Southern analysis confirmed that the cloned DNA was derived from *B. stearothermophilus* XL-65-6. This probe hybridized to a single band in each digestion of XL-65-6 genomic DNA but did not hybridize to *E. coli* DNA (Fig. 10). These results are consistent with the presence of a single chromosomal copy.

DNA Sequence Analysis and Gene Identification

The 5,211 bp region in pLOI1801 was sequenced in both directions by using a combination of nested deletions and overlapping subclones (Fig. 11). Five open reading frames (ORFs) were found in the sense strand, four of which were complete. All appear to be transcribed opposite to that of the *lac* promoter on the vector (Fig. 9).

The central region of plasmid pLOI1801 contains *ptsH*, *ptsI* and *ptsT*. The flanking ORFs were provisionally identified by amino acid homology to known genes from *B. subtilis* and corresponding gene designations were retained. The translated 5' incomplete ORF (324 amino acids) is similar to the carboxyl-terminus of glucose-specific EII (*ptsG'*; 67% identity; 81% similarity) (Zagorec and Postma, 1992). The translated 3' flanking gene (254 amino acids) is similar to the amino-terminal end of a precursor for wall associated protein (*wapA*; 46% identity and 63% similarity) (Foster,

Figure 10. Southern hybridization analysis of *ptsHIT* operon. Lane assignment: Lane 1: λ marker; Lane 2: *pst*I digestion of *E. coli* DH5α genomic DNA; Lane 3 to Lane 7: digestions of *B. stearothermophilus* XL-65-6 genomic DNA with *Cla*I, *Eco*RI, *pst*I, *Sal*I, and *Sma*I, respectively.

1 2 3 4 5 6 7

Figure 11. Nucleotide sequence of *B. stearothermophilus* DNA fragment in pLOI1801. The deduced amino acid sequences for each ORF are placed below the first nucleotide of the corresponding codon. The putative promoter region is underlined and the -10 and -35 regions labeled. Regions complementary to UUUCCUCC, the 3' end of 16s rRNA from *B. stearothermophilus* (Van Charldorp et al., 1981), are underlined and labeled RB. Encoded proteins are labeled at their respective start codons. Stop codons are indicated by asterisks.

```
   1 GCACTTGTTGAACGTGAAAATCGGGATGACCTTCTCCGGCGGGGTCATTGACTTCTTGCT
       H  L  L  N  V  K  I  G  M  T  F  S  G  G  V  I  D  F  L  L
  61 GTTTGGCGTTCTGCCAAACCGGACGGCATGGTGGCTCGTCATTCCGGTCGGCTTGGTGTT
       F  G  V  L  P  N  R  T  A  W  W  L  V  I  P  V  G  L  V  F
 121 TGCCGTGATTTATTATTTCGGGTTCCGCTTTGCGATCCGCAAATGGGATTTGGCGACGCC
       A  V  I  Y  Y  F  G  F  R  F  A  I  R  K  W  D  L  A  T  P
 181 GGGCCGGGAAAAAACGGTCGAGGAAGCACCGAAAGCCGAGGCGGCCGCTGCTGGCGATCT
       G  R  E  K  T  V  E  E  A  P  K  A  E  A  A  A  A  G  D  L
 241 GCCGTACGAAGTGCTCGCCGCCCTTGGCGGAAAGGAAAACATCGAACATTTGGACGCATG
       P  Y  E  V  L  A  A  L  G  G  K  E  N  I  E  H  L  D  A  C
 301 CATTACGCGTTTGCGCGTTTCGGTCCATGATATCGGCCGGGTCGATAAAGACCGGTTGAA
       I  T  R  L  R  V  S  V  H  D  I  G  R  V  D  K  D  R  L  K
 361 AGCGCTCGGCGCCGCAGGGGTGCTTGAAGTCGGCAACAACGTGCAAGCGATTTTCGGTCC
       A  L  G  A  A  G  V  L  E  V  G  N  N  V  Q  A  I  F  G  P
 421 GAAATCGGATATGTTAAAAGGGCAAATTCAAGACATCATGCAAGGAAAAGCGCCGGCCCG
       K  S  D  M  L  K  G  Q  I  Q  D  I  M  Q  G  K  A  P  A  R
 481 GGCGGAAGAGAAACCGAAAACAGCGGCCTCGGAAGCCGCGGAGTCCGAAACGATCGCTTC
       A  E  E  K  P  K  T  A  A  S  E  A  A  E  S  E  T  I  A  S
 541 GCCGATGTCCGGAGAAATCGTTCCGCTCGCTGAAGTGCCGGACCAAGTGTTTTCGCAAAA
       P  M  S  G  E  I  V  P  L  A  E  V  P  D  Q  V  F  S  Q  K
 601 AATGATGGGCGATGGGTTCGCAGTCATGCCGACGGACGGCACGGTCGTCTCTCCGGTTGA
       M  M  G  D  G  F  A  V  M  P  T  D  G  T  V  V  S  P  V  D
 661 TGGGAAGATCATCAACGTCTTTCCGACGAAGCACGCCATTGGCATTCAGTCGGCCGGCGG
       G  K  I  I  N  V  F  P  T  K  H  A  I  G  I  Q  S  A  G  G
 721 GCATGAAATTTTGATCCACGTCGGCATCGATACGGTGAAGCTGAACGGCCAAGGGTTTGA
       H  E  I  L  I  H  V  G  I  D  T  V  K  L  N  G  Q  G  F  E
 781 AGCGCTTGTGAAGGAAGGCGACGAAGTGAAAAAAGGACAGCCGATTTTGCGCGTCGACCT
       A  L  V  K  E  G  D  E  V  K  K  G  Q  P  I  L  R  V  D  L
 841 TGATTATGTCAAACAGAACGCCCCATCGATCGTCACACCGGTGATTTTCACGAATCTCCA
       D  Y  V  K  Q  N  A  P  S  I  V  T  P  V  I  F  T  N  L  Q
 901 AGCCGGCGAAACGGTTCACGTCAACAAGCAAGGCCCGGTTGCTCGAGGAGAAGACGCTGT
       A  G  E  T  V  H  V  N  K  Q  G  P  V  A  R  G  E  D  A  V
 961 TGTGACGATCCGCTAAACTGCTTATTCGCGGCGGGGCGGGACCAAGCATTTTCCGTTGCC
       V  T  I  R  *                                           -35
1021 TCCCCGCCCCTGCATTTGATATGATGGCAAAGAAAAAATTGAAGAAAAGGGGAATGATGA
                          -10                            RS
1081 ACATGGCAGAAAAAACGTTTAAAGTCGTTTCTGATTCCGGCATCCACGCTCGTCCGGCGA
        M  A  E  K  T  F  K  V  V  S  D  S  G  I  H  A  R  P  A  T
        HPr →
1141 CGATTTTGGTGCAAACGGCGAGCAAATTCAACAGCGAAATCCAGCTTGAGTACAACGGCA
        I  L  V  Q  T  A  S  K  F  N  S  E  I  Q  L  E  Y  N  G  K
1201 AAACGGTGAACTTGAAATCGATCATGGGCGTTATGTCGTTAGGAATTCCGAAAGGGGCGA
        T  V  N  L  K  S  I  M  G  V  M  S  L  G  I  P  K  G  A  T
1261 CGATCAAAATCACGGCAGAAGGGGCCGATGCGGCAGAAGCAATGGCGGCGTTAACGGATA
        I  K  I  T  A  E  G  A  D  A  A  E  A  M  A  A  L  T  D  T
1321 CGTTGGCGAAAGAGGGTCTTGCAGAATAATGGGAAACGCAATCCGCGAAAAAACGATCCA
        L  A  K  E  G  L  A  E  *M  G  N  A  I  R  E  K  T  I  H
                 RS                EI →
1381 TGGGATTGCTGCATCGAGCGGTATTGCCATCGCGAAGGCATACCGCTTAGAAACCCCTGA
       G  I  A  A  S  S  G  I  A  I  A  K  A  Y  R  L  E  T  P  D
1441 TTTGGCAGCCGAAAAACGGACGGTCGCCGACGTCGAGGCGGAAATTGCGCGGCTTGAGGC
       L  A  A  E  K  R  T  V  A  D  V  E  A  E  I  A  R  L  E  A
```

```
1501 GGCGGTCGCGAAAGCGAAAGAAGAGCTGGAAGCCATCAAACAGCATGCCTTGGAAAAGCT
      A  V  A  K  A  K  E  E  L  E  A  I  K  Q  H  A  L  E  K  L
1561 TGGCGAAGACAAAGCTGCCATTTTTGCCGCCCACTTGCTTGTGCTTGACGACCCAGAATT
      G  E  D  K  A  A  I  F  A  A  H  L  L  V  L  D  D  P  E  L
1621 GCTGAACCCGATTAAAGAAAAAATCCAAACGGAGCGCGTCAATGCCGAGTACGCTCTCGA
      L  N  P  I  K  E  K  I  Q  T  E  R  V  N  A  E  Y  A  L  D
1681 TGAAACCGCCTTGTTTTTCATCTCGATGTTCGAGGCGATGGACAATGAATATATGAAAGA
      E  T  A  L  F  F  I  S  M  F  E  A  M  D  N  E  Y  M  K  E
1741 ACGGGCCGCTGACATCCGCGATGTGACAAAGCGCGTCCTCGCCCATCTGCTTGGCGTCAC
      R  A  A  D  I  R  D  V  T  K  R  V  L  A  H  L  L  G  V  T
1801 AATCTCGAACCCGAGCCTCATTTCTGAGGAAGTCGTGATCATCGCGGAAGACTTGACGCC
      I  S  N  P  S  L  I  S  E  E  V  V  I  I  A  E  D  L  T  P
1861 ATCCGATACGGCGCAGCTGAACCGCCAATATGTGAAAGGATTTGCCACCGACATCGGCGG
      S  D  T  A  Q  L  N  R  Q  Y  V  K  G  F  A  T  D  I  G  G
1921 GCGGACGTCGCATTCAGCAATTATGGCCCGTTCGCTCGAAATCCCTGCCGTCGTCGGCAC
      R  T  S  H  S  A  I  M  A  R  S  L  E  I  P  A  V  V  G  T
1981 GAAGACGGTGACGGCGGAAGTAAAAAACGGCGACATCGTCATTGTCGATGGGCTCGACGG
      K  T  V  T  A  E  V  K  N  G  D  I  V  I  V  D  G  L  D  G
2041 CCAAGTAATCATCAATCCATCGCCGGAGTTGCTCGCCCAATATGAACAAAAACGGGCGCG
      Q  V  I  I  N  P  S  P  E  L  L  A  Q  Y  E  Q  K  R  A  R
2101 CTACGAGGCGCAAAAGGCAGAATGGGCGAAACTCGTTCACGAGGCGACCGTGACGGCTGA
      Y  E  A  Q  K  A  E  W  A  K  L  V  H  E  A  T  V  T  A  D
2161 CGGCATTCATGTCGAGCTGGCTGCCAACATCGGCACGCCAGACGATGTGAAAGGAGCGTT
      G  I  H  V  E  L  A  A  N  I  G  T  P  D  D  V  K  G  A  L
2221 GGCGAACGGAGCGGAAGGAATCGGATTGTATCGCACAGAATTTCTATACATGGGACGCTC
      A  N  G  A  E  G  I  G  L  Y  R  T  E  F  L  Y  M  G  R  S
2281 GGAACTGCCGACGGAAGACGAACAGTTTGTGGCTTACAAAACGGTGCTGGAACAAATGAA
      E  L  P  T  E  D  E  Q  F  V  A  Y  K  T  V  L  E  Q  M  N
2341 TGGCAAGCCGGTCGTTGTGCGGACGCTTGACATTGGCGGCGACAAAGAGCTGCCGTATTT
      G  K  P  V  V  V  R  T  L  D  I  G  G  D  K  E  L  P  Y  L
2401 GCAGCTGCCGAAAGAAATGAATCCGTTCTTAGGGTTCCGGGCAATCCGCCTTTGCCTTGA
      Q  L  P  K  E  M  N  P  F  L  G  F  R  A  I  R  L  C  L  E
2461 GATGCAAGATATGTTCCGCACGCAGCTGCGCGCCTTGTTGCGGGCGAGTGTGTACGGCAA
      M  Q  D  M  F  R  T  Q  L  R  A  L  L  R  A  S  V  Y  G  N
2521 TTTGAAAATCATGTTCCCGATGATTGCGACGCTCGATGAATTCCGCCAAGCGAAAGCCAT
      L  K  I  M  F  P  M  I  A  T  L  D  E  F  R  Q  A  K  A  I
2581 TTTGCTTGAAGAAAAAGAGGCGCTCCTCCGCCAAGGGGTTGCCGTTGCGGATGGAATTGA
      L  L  E  E  K  E  A  L  L  R  Q  G  V  A  V  A  D  G  I  E
2641 AGTCGGCATGATGGTGGAAATTCCGGCTGCCGCCGTCATGGCGGACCAGTTTGCGAAAGA
      V  G  M  M  V  E  I  P  A  A  A  V  M  A  D  Q  F  A  K  E
2701 AGTCGATTTCTTCAGCATTGGAACGAACGACCTGATCCAATATACGATGGCGGCCGATCG
      V  D  F  F  S  I  G  T  N  D  L  I  Q  Y  T  M  A  A  D  R
2761 GATGAATGAGCGGGTTGCGTATTTGTACCAGCCGTACAACCCGGCTATTTTGCGGCTCAT
      M  N  E  R  V  A  Y  L  Y  Q  P  Y  N  P  A  I  L  R  L  I
2821 CAGCCACGTGATCGATGCCGCCCATCGCGAAGGAAAATGGGTCGGGATGTGCGGGGAAAT
      S  H  V  I  D  A  A  H  R  E  G  K  W  V  G  M  C  G  E  M
2881 GGCCGGCGACCCGATCGCGATTCCGATTTTGCTTGCTTTAGGCCTTGATGAGTTCAGCAT
      A  G  D  P  I  A  I  P  I  L  L  A  L  G  L  D  E  F  S  M
2941 GAGCGCCACTTCGATTTTGCCGGCGCGCGCCCAGTTAAAAAAGCTGGCGAAAGAAGAGGC
      S  A  T  S  I  L  P  A  R  A  Q  L  K  K  L  A  K  E  E  A
3001 GGCCCGCATCAAAGAGACGGTGCTGTCGCTCGGCACGGCTGAGGAAGTCGTGTCGTTTGT
      A  R  I  K  E  T  V  L  S  L  G  T  A  E  E  V  V  S  F  V
```

```
3061 CAAACGAACATTTTCACTTGCCTGAATGAAAGCGTGGCGACGTTTACGTTCGCCGCGCTT
      K  R  T  F  S  L  A  *  M  K  A  W  R  R  L  R  S  P  R  F
                              PtsT →
3121 TTTTCGTTTTGCGTTATGCCAACAGCGGCGTCGACGATTCATAGCCATACCGCTGCCGGA
      F  R  F  A  L  C  Q  Q  R  R  R  R  F  I  A  I  P  L  P  D
3181 CCGTAAGCGAAAGGCCGATAGCGATAACTTTCGCCATTTGGTAAACAATGTGCAAATTCG
      R  K  R  K  A  D  S  D  N  F  R  H  L  V  N  N  V  Q  I  R
3241 TATACGGAAGCACGCTCATATCAGCCGTGTCCTCCATCATGACTCCCATGACGCTGATGT
      I  R  K  H  A  H  I  S  R  V  L  H  H  D  S  H  D  A  D  V
3301 CTCCGACCGGTGGCAGCGATTTGCCGAGCGCCGTGCCAGGGGAAAGGGCGCCAGGGCGGA
      S  D  R  W  Q  R  F  A  E  R  R  A  R  G  K  G  A  R  A  D
3361 TGGCGATCGTATGGACAAACGGCTGCGGCCCGACAATGCTGTCGATTGCCACTACATATG
      G  D  R  M  D  K  R  L  R  P  D  N  A  V  D  C  H  Y  I  C
3421 CTCCGTGTTCGTTTTGCAATTGTTCGGAAACGCGTCGGATGTTTGTCGCATCGACGGGTT
      S  V  F  V  L  Q  L  F  G  N  A  S  D  V  C  R  I  D  G  L
3481 GCCTGAGCGTGCCGATGACGGTCAAATGATCCGGATAAGCGTTTTCCAGCAGCGTGCCGA
      P  E  R  A  D  D  G  Q  M  I  R  I  S  V  F  Q  Q  R  A  D
3541 CGAACGGCCCTAAGCTGTCGCCGTTGATGCGATTGCTGCCGATGCAAACGACGGTCAGAT
      E  R  P  *
3601 GACGGATGTGCGCCGGCAGCAAGGAAAACAGTTCATTGCGAATGAGCAAAGGAGCGAGCG
3661 GATCGCTGTAGGAAATCGTCCTCATCAATCGCGATTCCTTTCTGTCTATAGTCTCCTTTT
3721 CTTATTCGCGGCCGCTGCCGGCGATTCCTTTCTTTTCAGGGTTGCAAATTTTTCCTTATG
3781 TTTGCTTAAACGGGTTGCGCCATTCCATTAACACGGCATGGTTGAGCGATTCTTCGTCTT
3841 CTAAACTTTCTTGCTGCACGCGGATGAGATCAGCAAAGTCGTCTGTCGTATAGTTGCGGA
3901 TGACCGCTTCGCGCGGGGTGTTGCCGTCAAAGACGTATAAACGTTTATCGTACATGCTGA
3961 TCCCCCTTTCGTTCGTTTTACCCTCATCATAACCGAAACCTAGAAAAAGCGCCAAGAAAG
4021 GTCCATGGACAAGTTTGCCCTCGACCCGTCCATGGACAAGTTTGCCCTCGACCCGGGGAT
4081 TCACGGAAAGCGGGTAGCTTGGTGTTCAAATAAAACGATCATTTACAGTCTTACGCTAAG
4141 TGATCAAGATCACCATCGCAAGCAACCACTGGCAATCTTCGTTTAGGACTCTTTCTCCAT
4201 TTCACATACAAATAGGTTTACTCGCAAAAGCTTCCGAAGAAAATGAAAAACCATGTCGAA
4261 CTTTGTCTGTCGAAAATGATGGATGGCAGGGAGCAGGAAAGATACAATTTGATTGTAAAA
4321 GTAATGGAAATCGGGAAGGAGAGCATATGAGAAACATTCGTTTGTGATGATGGTTGGTTT
4381 GGATGGTCGTATGGGCTTTAGTGATGCCAAGCGTCCCGACAAAGGGACTTGCCGAAACGA
       M  V  V  W  A  L  V  M  P  S  V  P  T  K  G  L  A  E  T  I
       WapA' →
4441 TCCGATCAGGGATTGACCGTGTAGAGCCAAAAGAGGAGCCGCCAAAGGTGCCGCAAGCCC
       R  S  G  I  D  R  V  E  P  K  E  E  P  P  K  V  P  Q  A  P
4501 CGAAACGTGATCTCAAGCCAGGGGAAATCATAGAAGAAAGAACTGAGAATACAAAGGTTT
       K  R  D  L  K  P  G  E  I  I  E  E  R  T  E  N  T  K  V  Y
4561 ACTACAACGGCGATGGCACGTTTACCAAGAAGATTTACTTTGAGCCTATTCATGTGAAGA
       Y  N  G  D  G  T  F  T  K  K  I  Y  F  E  P  I  H  V  K  K
4621 AAAAGGGACAAAAAATATTTGAGGAAGTATCTTCTAGCCTAACGGATAGCACCAATAATA
       K  G  Q  K  I  F  E  E  V  S  S  S  L  T  D  S  T  N  N  T
4681 CGAATTACGTTGAAACAGAGAATACCATTTTGGAGACAAACTTTTATAAAAAGATGGTCG
       N  Y  V  E  T  E  N  T  I  L  E  T  N  F  Y  K  K  M  V  D
4741 ATGGAGAATACGCCAATTTCCATTACAATGGATATTCGATTTCCTATTCTATTTTAGAGG
       G  E  Y  A  N  F  H  Y  N  G  Y  S  I  S  Y  S  I  L  E  A
4801 CGGCAGGTAATGACGTACAATCGATAAAAGCGAAGGATGTTGCCGCCGTCTATAAGAAAA
       A  G  N  D  V  Q  S  I  K  A  K  D  V  A  A  V  Y  K  K  K
4861 AAGACAACAAAATTCTTCACAAAAACATTTTTCCGAGCATAGACCTACAAAATATTACAT
       D  N  K  I  L  H  K  N  I  F  P  S  I  D  L  Q  N  I  T  F
4921 TTAACGAGTCAACAAAAGAAGATTTGGTTCTTCATTCGGTTGGCTATCATATTTTCAAAT
```

```
       N  E  S  T  K  E  D  L  V  L  H  S  V  G  Y  H  I  F  K  F
4981 TCCGCCTGAAAACGGATTTGCAAGCAGACATTCAAGACGATGGCTCCATTTTATTGACAA
       R  L  K  T  D  L  Q  A  D  I  Q  D  D  G  S  I  L  L  T  N
5041 ATCAAGAACATGAAAAGGTGTTTGAACTGCCGAAGCCGTTTATGGTGGATTCGAATGTCG
       Q  E  H  E  K  V  F  E  L  P  K  P  F  M  V  D  S  N  V  D
5101 ATGAACACTCTGGTGAAGTGCAGCGTCTCAGAGAATGTAACGTATGAATTACAACAGGAT
       E  H  S  G  E  V  Q  R  L  R  E  C  N  V  *
5161 ATCAAGGATATATTTTAACGGTCAAAGCCGACCCTGATTGGCTGAAGGATC
```

1993). Although stem loop structures can be predicted between *ptsG* and *ptsH* and between *ptsT* and *wapA'*, none resembles a *rho*-independent terminator.

In the anti-sense strand, 6 ORFs were found which ranged from 140 to 317 codons in length. Three were very interesting. The largest ORF extended from base 434 to 1381, spanning from *ptsI* to *ptsG'*. Part of the translated amino acid sequence for this ORF is homologous to *B. subtilis* protein K (Gonzy-Tréboul et al., 1989) which is also located on the complementary strand of *ptsH* and has been expressed *in vivo*. Another interesting ORF extends from base 1270 to 2052, covering part of *ptsI* and 3' end of *ptsH* (262 codons). This ORF contains a high proportion of serine codons (19.2%). The translated sequence is similar to serine-rich proteins such as the α-agglutinin core subunit from *Saccharomyces cerevisiae* (Roy et al., 1991). The third ORF contains 184 codons and extends from base 3134 to 3685, covering part of *ptsT* and the downstream region. Part of this translated sequence is homologous to hypothetical protein 7 from *B. subtilis* (Ogasawara and Yoshikawa, 1992). No highly conserved ribosomal-binding sites were found immediately upstream from these three ORFs.

The *ptsHIT* Operon

The three complete ORFs in the central region of the cloned fragment appear to comprise an operon, *ptsHIT*. A putative promoter is located 36 bases upstream of the start codon for *ptsH*. This region resembles the consensus sequence for *B. subtilis* σ^A, a promoter which is primarily associated with constitutive housekeeping genes (Moran, 1993). The sequence 11 bp upstream from the *ptsH* coding region closely resembles the 3' end of 16*s* ribosomal RNA from *B. stearothermophilus* (Van Charldorp et al., 1981) and is nearly identical to the proposed ribosomal-binding sites for two PTS genes from *B. subtilis, ptsH* (Gonzy-Tréboul et al., 1989) and *bgl* (Murphy et al., 1984). A potential ribosomal-binding site for *ptsI* is located within the carboxyl-terminus of *ptsH*. The ATG start codon for *ptsI* overlaps the TAA stop codon for *ptsH* as had been previously reported for the *ptsHI* operon in *B. subtilis* (Gonzy-Tréboul, 1989). A third ORF designated *ptsT* begins immediately adjacent to the stop codon of *ptsI*. No ribosomal-binding site was readily identified for *ptsT*.

The *ptsH* gene encodes a peptide containing 88 amino acids with a calculated M_r of 9,166. This translated sequence is more similar to HPrs from other Gram positive bacteria (70.5% to 64.8% identity) than to HPrs from Gram negative bacteria (42.0% to 31.8%) (Table 12). Three highly conserved regions were found which are identical in all Gram positive HPrs in the current GenBank data base (Fig. 12A). The first spans Gly^{13} to Thr^{20} and includes His^{15} which serves as the phosphoryl acceptor for EI (Weigel et al., 1982). The second extends from Leu^{22} to Lys^{28}. The third contains 13 amino acids (Val^{42} to Gly^{54}) and includes Ser^{46}, a site which can be phosphorylated by an ATP-dependent kinase in other Gram positive bacteria (Postma et al., 1993).

Patterns of codon usage are remarkably different for *B. stearothermophilus* and *B. subtilis* HPrs although the amino acid

Table 12. Comparison of predicted amino acid sequences of *ptsH* and *ptsI* from *B. stearothermophilus* to those from other bacteria

Proteins	% Identity (% Similarity) *Bst*-HPr	 *Bst*-EI	Accession Number[a]
HPr			
Gram positive organisms			
B. subtilis	70.5(84.1)		X12832
E. faecalis	68.2(79.5)		A25053
S. mutans	67.8(81.6)		L15191
S. salivarius	65.5(79.3)		Z17217
S. carnosus	64.8(80.7)		X60766
S. aureus	62.1(78.2)		P02907
Gram negative organisms			
M. capricolum	42.0(60.2)		L22432
A. eutrophus	36.8(59.8)		M69036
E. coli	32.9(48.2)		J02796
S. typhimurium	32.9(48.2)		M21450
K. pneumoniae	31.8(48.2)		X51452
Enzyme I			
Gram positive organisms			
B. subtilis		76.5(87.4)	M98359
S. carnosus		63.7(77.0)	M69050
S. mutans		63.7(76.8)	L15191
S. salivarius		63.5(76.4)	M81756
Gram negative organisms			
S. typhimurium		50.3(69.8)	M76176
E. coli		49.9(69.5)	J02796
A. eutrophus		40.0(61.6)	M69036

[a] GenBank accession number

Figure 12. Comparison of deduced amino acid sequences of *ptsH* and *ptsI* from *B. stearothermophilus* and from other Gram positive bacteria. Regions identical in HPrs from 4 or more organisms are indicated by shadowing. Regions identical in EIs from 3 or more organisms are indicated by shadowing. Gaps (dashes) have been introduced to optimize alignment.

A. HPr

```
Bst MAEKTFKVVSDSGIHARPATILVQTASKFNSEIQLEYNGKTVNLKSIMGVMSLGIPKGAT
Bsu MAQKTFKVTADSGIHARPATVLVQTASKYDADVNLEYNGKTVNLKSIMGVMSLGIAKGAE
Efa MEKKEFHIVAETGIHARPATLLVQTASKFNSDINLEYKGKSVNLKSIMGVMSLGVGQGSD
Smu MASKDFHIVAETGIHARPATLLVQTASKFASDITLDYKGKAVNLKSIMGVMSLGVGQGAD
Ssa MASKDFHIVAETGIHARPATLLVQTASKFASDITLDYKGKAVNLKSIMGVMSLGVGQGAD
Sca MEQQSYTIIDETGIHARPATMLVQTASKFDSDIQLEYNGKKVNLKSIMGVMSLGVGKDAE
Sau MEQNSYVIIDETGIHARPATMLVQTASKFDSDIQLEYNGKKVNLKSIMGVMSLGVGKDAE

Bst IKITAEGADAAEAMAALTDTLAKEGLAE-
Bsu ITISASGADENDALNALEETMKSEGLGE-
Efa VTITVDGADEAEGMAAIVETLQKEGLAEQ
Smu VTITAEGADADDAIAAINETMTKEGLA--
Ssa VTISAEGADADDAIVAIAETMTKEGLA--
Sca ITIYADGSDEADAIQAITDVLSKEGLTE-
Sau ITIYADGSDESDAIQAISDVLSKEGLT--
```

B. EI

```
Bst MGNAIREKTIHGIAASSGIAIAKAYRLETPDLAAEKRTVADVEAEIARLEAAVAKAKEEL
Bsu M------QELKGIGASAGIAIAKAYRLEEPDLTVEKKNISDSEAEVSRFDEAIARSKEEL
Sca MA-----KQIKGIAASDGVAIAKAYLLVEPDLSFDNESVTDTDAEVAKFNGALNKSKVEL
Smu MTEM-----LKGIAASDGVAVAKAYLLVQPDLTFETVSVTDTQAEEARLDAALEASQNEL
Ssa MTEM-----LKGIAASDGVAVAKAYLLVQPDLSFETVTVEDTSAEEARLDAALAASQDEL

Bst EAIKQHALEKLGEDKAAIFAAHLLVLDDPELLNPIKEKIQTERVNAEYALDETALFFISM
Bsu EKIKEHALKELGQ--------------------PVKEKISTDSVNAEFALKETSSMFVTM
Sca TKIRNNAEKQLGADKAAIFDAHLLVLEDPELIQPIEDKIKNESVNAAQALTDVSNQFITI
Smu SLIRQKAVDTLGEEAAAVFDAHLMVLADPEMIGQIKETIRTKEVNAESALKEVTDMFVTL
Ssa SVIREKAVESLGEEAAAVFDAHLMVLADPEMTGQIKETIRAKQVNAEAALTEVTDMFIAI

Bst FEAMDNE-YMKERA-ADIRDVTKRVLAHLLGVTISNPSLISEEVVIIAEDLTPSDTAQLN
Bsu FESMDNE-YMKERA-ADIRDVTKRVTGHLLGVEIPNPSMISEEVIIVAEDLTPSDTAQLN
Sca FESMDNE-YIAERAKADIRDVSKRVLAHILGVELPNPSIVDESVVIIGNDLTPSDTAQLN
Smu FENMEDNPYMQERA-ADIRDVAKRVLAHLLGVELPNPATISEESIVIAHDLTPSDTAQLD
Ssa FEGMEDNPYMQERA-ADIRDVTKRVLANLLGKKLPNPATINEESIVVAHDLTPSDTAQLD

Bst RQYVKGFATDIGGRTSHSAIMARSLEIPAVVGTKTVTAEVKNGDIVIVDGLDGQVIINPS
Bsu REFVKGFTTDIGGRTSHSAIMARSLEIPAVVGTKAATGTIQNGVTVIVDGINGDVIIDPS
Sca KEYVQGFVTNIGGRTSHSAIMSRSLEIPAVVGTKSITEEVEAGDTIVVDGMTGDVLINPS
Smu ANYVKAFVTNIGGRTSHSAIMARTLEIAAVLGTNDITERVKNGDIVAVNGITGQVIINPT
Ssa KKYVKAFVTNIGGRTSHSAIMARTLEIAAVLGTNNITELVKDGDILAVSGITGEVVINPT

Bst AELLAQYEQKRARYEAQKAEWAKLVHEATVTADGIHVELAANIGTPDDVKGALANGAEGI
Bsu AETVKEYEEKHNAYLAQKAEWAKLVNEPTVSKDGHHVELAANIGTPDDVKGVLENGGEAV
Sca DEVIAEYQEKRENFFKDKQELQKLRDAESVTADGHHVELAANIGTPNDLPGVIENGAEGI
Smu EDQIAEFKAAGETYAKQKAEWALLKDAETVTADGKHFELAANIGTPKDVEGVNNNGAEAV
Ssa EEQIAEFKAAGEAYAKQKAEWALLKDAQTVTADGKHFELAANIGTPKDVEGVNDNGAEAV
```

```
Bst GLYRTEFLYMGRSELPTEDEQFVAYKTVLEQMNGKPVVVRTLDIGGDKELPYLQLPKEMN
Bsu GLYRTEFLYMGRDQLPTEDEQFDAYKTVLERMEGKSVVVRTLDIGGDKELPYLQLPKEMN
Sca GLYRTEFLYMGRDQMPTEEEQFEAYKAVLEAMKGKRVVVRTLDIGGDKELPYLDLPEEMN
Smu GLYRTEFLYMDSQDFPTEDEQYEAYKAVLEGMNGKPVVVRTMDIGGDKELPYFDLPKEMN
Ssa GLYRTEFLYMDSQDFPTEDDQYEAYKAVLEGMNGKPVVVRTMDIGGDKELPYFDLPKEMN

Bst PFLGFRAIRLCLEM--QDMFRTQLRALLRASVYGNLKIMFPMIATLDEFRQAKAILLEEK
Bsu PFLGYRAIRLCLEE--QEIFRTQLRALLRASTYGNLKIMFPMIATVNEFKEAKAILLEEK
Sca PFLGYRAIRLCLDQ--PEIFRPQLRALLRASVFGKLNIMFPMVATIQEFRDAKAFLKKNV
Smu PFLGFRALRISISETGNQMFRTQLRALLRASVHGQLRIMFPMVALLNEFRKAKGILEEEK
Ssa PFLGYRALRISISETGNQMFRTQLRALLRASVHGKLRIMFPMVALLTEFRTAKGILEEEK

Bst EALLRQGVAVADGIEVGMMVEIPAAAVMADQFAKEVDFFSIGTNDLIQYTMAADRMNERV
Bsu EKLVKAGQAVSDDIEVGMMVEIPSTAVIADQFAKEVDFFSIGTNDLIQYTMAADRMNERV
Sca LTLKMKAMKVADDIELGIMVEIPSTAALADIFAKEVDFFSIGTNDLIQYTMAADRMSERV
Smu ANLKAEGVAVSDDIQVGIMIEIPAAAMLADQFAKEVDFFSIGTNDLIQYTMAADRMNEQV
Ssa AKLVAEGVAVADDIEVGIMIEIPAAAMLADQFAKEVDFFSIGTNDLIQYTMAADRMNEQV

Bst AYLYQPYNPAILRLISHVIDAAHREGKWVGMCGEMAGDPIAIPILLALGLDEFSMSATSI
Bsu SYLYQPYNPAILRLITLVIEAAHKEGKWVGMCGEMAGDEIAIPILLGLGLDEFSMSATSI
Sca SYLYQPYISNFSFLVKQVIEASHAEGKWTGMCGEMAGDQTAIPLLLGLGLDEFSMSATSI
Smu SYLYQPYNPSILRLVDHVVKAAHAEGKWAGMCGEMAGDQTAVPLLVGIGLDEFSMSATSV
Ssa SYLYQPYNPSILRLINNVIKAAHAEGKWAGMCGEMAGDQTAVPLLVGMGLDEFSMSATSV

Bst LPARAQLKKLAKEEAARIKETVLS-LGTAEEVVSFVKRTFSLA
Bsu LPARTQISKLSKQEAESFKEKILS-MSTTEEVVAFVKETFK--
Sca LKARVLIRSLNESEMKELSERAVQ-CATSEEVVDLVEEYTKNA
Smu LRTRSLMKKLDTAKMQELAQRALTECATMEEVLELEKEYIDFD
Ssa LRTRSLMKKLDTAKMEEYANRALTECSTMEEVLELSKEYVNVD
```

sequences are very similar (70% identity). Of 28 identical amino acids in the three perfectly conserved regions, 15 are translated from different codons resulting from 17 base changes. The replacement of A and T in *B. subtilis* with G or C in *B. stearothermophilus* accounts for 82% of these changes. However, base changes for the remaining aligned amino acids appear to be quite random with a 48% frequency of replacement of A and T for G and C (Fig. 13).

The *ptsI* gene encodes 578 amino acids (M_r 63,470). Although this gene encodes a protein which is very similar to enzyme I from *B. subtilis* (Table 12 and Fig. 12B), the translated sequence for *B. stearothermophilus ptsI* contains considerably less serine and threonine with correspondingly higher levels of alanine and leucine. As with *ptsH*, the translated amino acid sequence of *ptsI* from *B. stearothermophilus* is more similar to Gram positive bacteria (76.5% to 63.5% identity) than to Gram negative bacteria (50.3% to 40.0% identity) (Table 12). Highly conserved regions in EI are scattered within the middle of the polypeptide, the longest being 25 amino acids. Both amino- and carboxyl-terminal regions are less conserved (Fig. 12B). It is interesting to note that the *B. stearothermophilus* EI also shares partial homology with pyruvate phosphate dikinases from eukaryotes and prokaryotes (Bruchhaus and Tannich, 1993; Pocalyko et al., 1990) and with *E. coli* PEP synthase (Niersbach et al., 1992).

The *ptsT* gene potentially encodes 155 amino acids. No

Figure 13. Comparison of nucleotide sequences of *ptsH* from *B. stearothermophilus (B. st)* and from other Gram positive bacteria (*B. su: B. subtilis; E. fa: E. faecalis; S. mu: S. mutans; S. sa: S. salivarius*). Identical residues are indicated by asterisks. The deduced amino acid sequences in three identical regions are listed below the first nucleotide of the corresponding codon. The phosphorylation sites are overlined. Gaps (dashes) have been introduced to optimize alignment.

```
B. st    ATG GCA GAA AAA ACG TTT AAA GTC GTT TCT GAT TCC GGC ATC CAC GCT CGT CCG GCG ACG
B. su    ATG GCA CAA AAA ACA TTT AAA GTA ACT GCA GAT TCT GGA ATC CAT GCT CGT CCT GCG ACA
E. fa    ATG GAA AAG AAA GAA TTT CAC ATT GTA GCA GAA ACA GGA ATT CAC GCA CGT CCA GCT ACT
S. mu    ATG GCT TCA AAA GAT TTT CAC ATT GTT GCA GAA ACA GGA ATT CAT GCA CGC CCA GCT ACT
S. sa    ATG GCT TCT AAA GAT TTC CAC ATT GTT GCA GAA ACA GGT ATC CAC GCA CGT CCA GCT ACT
         *** *       ***     **   *   *       *  **   *  **  **  **  **  **  **  **  **
                                                         G   I   H   A   R   P   A   T

B. st    ATT TTG GTG CAA ACG GCG AGC AAA TTC AAC AGC GAA ATC CAG CTT GAG TAC AAC GGC AAA
B. su    GTT CTT GTA CAA ACT GCT AGC AAA TAC GAC GCT GAC GTT AAT TTA GAA TAT AAC GGC AAA
E. fa    TTA TTA GTA CAA ACT GCA AGC AAA TTT AAC TCA GAT ATT AAC TTA GAA TAC AAA GGT AAA
S. mu    TTG CTT GTT CAA ACA GCA AGT AAG TTT GCC TCA GAT ATT ACA CTT GAT TAC AAA GGA AAA
S. sa    TTG CTT GTT CAA ACA GCT AGC AAA TTT GCT TCA GAT ATC ACT CTT GAC TAC AAA GGT AAA
          *   *  **  *** **  **  **  **  *           **   *       *  **  **  **  **  ***
         .   L   V   Q   T   A   S   K

B. st    ACG GTG AAC TTG AAA TCG ATC ATG GGC GTT ATG TCG TTA GGA ATT CCG AAA GGG GCG ACG
B. su    ACA GTT AAC CTT AAA TCT ATT ATG GGT GTT ATG TCT TTA GGT ATC GCT AAA GGC GCT GAG
E. fa    TCT GTT AAC TTA AAA TCA ATC ATG GGC GTT ATG TCT TTA GGC GTT GGT CAA GGT TCT GAC
S. mu    GCA GTA AAC CTT AAG TCA ATT ATG GGT GTT ATG AGC CTT GGT GTT GGT CAA GGT GCT GAT
S. sa    GCA GTT AAC CTT AAA TCT ATC ATG GGT GTT ATG AGC CTT GGT GTT GGT CAA GGT GCT GAC
             **  ***  *  **  **  **  *** **  *** ***      *  **   *       ** *    *
             V   N   L   K   S   I   M   G   V   M   S   L   G

B. st    ATC AAA ATC ACG GCA GAA GGG GCC GAT GCG GCA GAA GCA ATG GCG GCG TTA ACG GAT ACG
B. su    ATC ACA ATC TCT GCT TCC GGA GCT GAC GAA AAC GAT GCT CTT AAC GCT TTA GAA GAA ACA
E. fa    GTA ACA ATC ACT GTT GAT GGT GCT GAC GAA GCT GAA GGA ATG GCA GCA ATC GTT GAA ACA
S. mu    GTT ACA ATC ACT GCT GAA GGT GCA GAT GCA GAT GAT GCA ATT GCA GCT ATT AAT GAA ACG
S. sa    GTA ACT ATC TCT GCA GAA GGT GCT GAT GCT GAT GAC GCA ATC GTA GCA ATC GCA GAA ACA
          *  *   ***     *       **  **  **  *       **  *    *      **   *      **  **

B. st    TTG GCG AAA GAG GGT CTT GCA GAA TAA
B. su    ATG AAA AGC GAA GGA CTC GGC GAG TAA
E. fa    TTA CAA AAA GAA GGA TTG GCT GAA TAA
S. mu    ATG ACT AAA GAA GGA TTG GCT --- TAA
S. sa    ATG ACA AAA GAA GGA TTG GCA --- TAA
          *      *   **  **   *          ***
```

significant matches to the translated sequence for this protein were found in the current GenBank data base.

General characteristics of the *ptsH, ptsI*, and *ptsT* products are listed in Table 13. The isoelectric points for HPr and EI were 5.9 and 4.9, respectively, higher than reported for other Gram positive organisms which had pI values ranging from pH 4.15 to pH 4.6 for HPrs and from pH 4.40 to pH 4.63 for EIs. Both *B. stearothermophilus* HPr and EI had fewer negatively charged amino acid residues. The predicted pI for the *ptsT* product was surprisingly high, pH 11.7. The translated *ptsT* gene contained a high proportion of basic amino acids (27%), Arg alone comprising 19%.

Temperature-Induced Hyperexpression of *ptsHI* in *E. coli*

Expression of the *ptsHIT* operon in *E. coli* JLT2 (*ptsI* mutant) was examined in cells grown in Luria broth at 30°C and 37°C using denaturing gels (Fig. 14). Two new protein bands were observed in separated extracts of JLT2(pLOI1801) and JLT2(pLOI1804) which correspond in size to HPr and EI, apparent M_r of 10,000 and 64,000, respectively (Fig. 14). No new bands were observed which corresponded to the *pstT* although expression at a modest level could easily be obscured by native proteins of similar size. With both pLOI1801 and pLOI1804, HPr and EI levels were estimated to be at least 5-fold higher in cells grown at 37°C than in cells grown at 30°C.

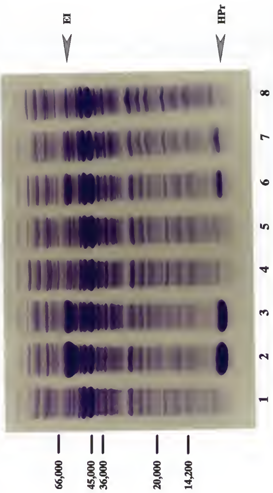
Polar Mutation in *ptsH*

A *Sac*I linker (pCGAGCTCG) was inserted into plasmid pLOI1806

Table 13. Characteristics of *ptsHIT* operon from *B. stearothermophilus*

Proteins	No. of nucleotides	%GC	Total No. of amino acids	M_r	Charge	Predicted pI
HPr	267	50.2	88	9,166	−1	5.9
EI	1737	54.7	578	63,470	−23	4.9
PTST	468	55.8	155	18,428	+15	11.7

Figure 14. SDS-PAGE of protein extracts from JLT2 recombinants. Approximately 25 μg of protein was loaded in each lane. Positions of proteins used as size markers are noted on the left. Arrowheads on the right indicate positions of HPr and EI. Lanes 1 through 4 contain extracts from cells grown at 30°C and lanes 5 through 8 contain extracts from parallel cultures grown at 37°C. Lanes: 1&5, JLT2(pUC18); 2&6, JLT2(pLOI1804); 3&7, JLT2(pLOI1801); 4&8, JLT2(pLOI1807).


EI
HPr
66,000
45,000
36,000
20,000
14,200
1
2
3
4
5
6
7
8

(Fig. 9B) at the *Hpa*I site of *ptsH* to create a frame-shift mutation (pLOI1807). This frame-shift mutation replaced the carboxyl-terminal 12 amino acids with a new 45 amino acid segment. After transformation into 1101 (*ptsH* mutant) and MM6 (*ptsI* mutant), recombinants were tested on carbohydrate-MacConkey agar plates. Strain 1101(pLOI1807) was negative, as expected. MM6(pLOI1807) exhibited weak enzyme I activity at 37°C (light pink colonies) but was negative at 30°C indicating that functional expression of *ptsI* was greatly reduced or eliminated by the upstream frame-shift mutation in *ptsH*. However, no protein bands corresponding to HPr or EI were observed following SDS-PAGE of extracts from recombinants containing pLOI1807 (Fig. 14).

Growth Inhibition of *E. coli ptsI* Mutants by *B. stearothermophilus ptsHIT* Operon

In the initial screening, toxicity problems associated with the *B. stearothermophilus* DNA containing *ptsHIT* prevented the recovery of positive clones by complementation of an *E. coli ptsI* mutant (strain MM6) at 37°C but not at 30°C. MM6 recombinants (pLOI1800, pLOI1801) harboring this operon grew well on PTS sugars at 30°C but formed small red colonies at 37°C. Although DH5α (*ptsI*$^+$) was not inhibited by the presence of plasmids containing *ptsHIT* at either temperature, two additional *E. coli ptsI* mutants (JLT2 and BL21) were also inhibited by this operon at 37°C. Toxicity of the *ptsHIT* operon at 37°C in *E. coli ptsI* mutants was not limited to a particular medium and did not require the addition of carbohydrate since similarly poor growth was observed with Luria agar lacking added sugar. This toxicity was eliminated by insertion of a frame shift mutation in *ptsH* (pLOI1807) which also decreased the expression of *ptsI* (Fig. 14). Toxicity of this operon was also eliminated by deleting the *ptsT* gene (pLOI1803 and pLOI1804) without affecting the ability of recombinants to produce high levels of functional EI and HPr at 30°C and 37°C (Fig. 14).

Construction and testing of plasmid pLOI1808 containing the complete cellobiose PTS system from *B. stearothermophilus* XL-65-6

The 3.8 kbp DNA fragment containing *B. stearothermophilus* XL-65-6 *cel* operon was isolated from pLOI902 (includes functional promoter) by digestion with *MluI* and *Nsi*I. This fragment was inserted between the *Pst*I (polylinker) and *Mlu*I sites of pLOI1803 which contains the XL-65-6 *ptsHI* genes (Fig. 9B). The resulting plasmid, pLOI1808, hydrolyzed cellobiose analogues (4-methylumbelliferyl-ß-D-glucopyranoside and 4-methylumbelliferyl-ß-D-cellobiopyranoside) and complemented *ptsH* and *ptsI* mutations in *E. coli*.

On MacConkey agar plates containing cellobiose and incubated at 37°C, DH5α(pLOI1808) formed pink colonies which were darker than transformants with either of the original plasmids (pLOI902 or pLOI1803). Controls with pUC18 alone formed light yellow zones due to an increase in pH from amino acid catabolism.

Attempts were made to improve the cellobiose utilization of DH5α(pLOI1808) by mutagenesis. Crystals of nitrosoguanidine were added to the center of cellobiose-MacConkey agar plates which had been spread with approximately 2,000 CFU of DH5α(pLOI1808). After incubation, sixteen colonies were recovered which appeared more intensely pink. Plasmids were purified from these colonies, transformed back into fresh DH5α, and recombinants tested. A mixture of colony types was obtained with each plasmid preparation. Most transformants appeared similar in color to MM6(pLOI1808) while others were white indicating loss of function. However, all contained a low level of intensely red colonies. None of these red colonies were stable during subculturing and none were able to grow on cellobiose-minimal agar. Further selection was attempted in minimal broth containing cellobiose but without success.

Discussion

The *ptsHIT* Operon

To our knowledge, this is the first time that *ptsH* and *ptsI* have been described from a thermophilic bacterium. In *B. stearothermophilus*, these two genes form an operon which includes a third gene, *ptsT*. The three genes in this operon may be translationally coupled because the last base of stop codon for *ptsH* is overlapped by the first base of start codon for *ptsI* and the stop codon for *ptsI* is adjacent to the start codon for *ptsT*. In recombinant *E. coli*, a mutation in the first gene dramatically decreased the expression of the downstream *ptsI* consistent with transcription as a single mRNA unit in *E. coli* and presumably also in *B. stearothermophilus*. Although no potential *rho*-independent terminators were found downstream from the *ptsHIT* operon, the long ORF in the antisense strand may serve this function in *B. stearothermophilus*. Antisense ORFs at the *ptsHI* locus from *E. coli* (Lévy et al., 1989) and from *B. subtlis* (Gonzy-Tréboul et al., 1989) have been expressed. Lévy et al. (1989) proposed that antisense transcription could mediate a premature termination of transcription at the *ptsH-ptsI* locus of *E. coli*.

The translated sequences for *ptsH* and *ptsI* from *B. stearothermophilus* shared extensive homology with the corresponding gene products from other bacteria. Though clearly related, homologues from gram positive bacteria were more similar to each other than to homologues from gram negative bacteria. Translated amino acid sequences for *B. stearothermophilus ptsH* and *ptsI* were most like those of *B. subtilis* but contained higher levels of alanine and leucine, with corresponding reductions in threonine and serine. This shift in amino acid composition may be related to thermal tolerance (Mozhaev and Martinek, 1984).

The translated sequence for the *ptsT* gene was not similar to any other entries in the current GenBank data base, although the size, high proportion of basic amino acids (Arg,Lys,His) and high pI resemble some bacterial histone-like proteins (Drlica and Rouviere-Yaniv, 1987). The *ptsT* gene was not essential for functional complementation of *E. coli* strains containing mutations in either *ptsH* or *ptsI* mutations. No third gene is present immediately adjacent to *ptsI* in the *B. subtilis* operon although a third ORF has been reported much further downstream (Reizer et al., 1993).

Toxicity Associated with the *ptsT* Gene

Clones containing the *B. stearothermophilus ptsHIT* operon were readily obtained from gene libraries by functional complementation of an *E. coli ptsI* mutant provided the initial selection was conducted at 30°C rather than 37°C. The *ptsT* gene was identified as the basis for toxicity at 37°C. Much higher levels of the *ptsH* and *ptsI* gene products were present in recombinant *E. coli* grown at 37°C than at 30°C. Deletion of the *ptsT* gene eliminated toxicity without decreasing expression of the recombinant *ptsH* and *ptsI* genes. A frame-shift mutation in *ptsH* which blocked the production of recombinant HPr also decreased the expression of *ptsI* (downstream) (and presumably *ptsT*) and eliminated toxicity.

Although the role of *ptsT* is unknown, it is interesting to note that a third gene of similar size is also present at the end of the operon containing *ptsH* and *ptsI* in *E. coli* and *Salmonella typhimurium*, the crr gene (Postma et al., 1993). The importance of the PTS and the crr gene product for catabolite repression in enteric bacteria is well established (Mitchell et al., 1989; Postma et al., 1993). Although the mechanisms differ, PTS components also regulate sugar utilization in gram positive bacteria (Postma et al., 1993). Since the toxicity of *B. stearothermophilus ptsT* gene was limited to *E. coli ptsI* mutants, *ptsT* products are presumed to interfere with the regulatory aspects involving PTS components in *E. coli*. It is tempting to speculate that *ptsT* may have some regulatory function in *B. stearothermophilus*.

Construction of a Recombinant Cellobiose PTS

Previous studies in our laboratory have attempted to develop a functional hybrid cellobiose uptake system in *E. coli* by combining the cellobiose-specific PTS genes (*celABCD*) from *B. stearothermophilus* with the *E. coli* general proteins (*pstHI*). Although function could be demonstrated with fluorescent analogues of cellobiose, the utilization of cellobiose was poor. To eliminate the potential problem of incompatibility between heterologous transport proteins, we constructed a plasmid which included the full set of *B. stearothermophilus* genes needed for cellobiose transport and cleavage. This combination of genes facilitated a low level of cellobiose utilization as indicated by acidification of cellobiose-MacConkey plates but was insufficient to support growth on cellobiose in minimal medium. Attempts to improve cellobiose utilization by plasmid mutagenesis were of limited success. Mutants appear to be selectable but are unstable. Clearly the *B. stearothermophilus* HPr and EI proteins functioned well with the native sugar specific PTS permeases in *E. coli*. Thus the inability of these recombinant *E. coli* to utilize cellobiose effectively may reside in the formation of an active EII^{Cel} complex. It is possible that effective assembly of the EII^{Cel} is hindered by differences in membrane composition or that higher temperatures are required for EII^{Cel} proper assembly and function.

CHAPTER 6
SUMMARY AND CONCLUSIONS

Endospore-forming thermophiles were readily isolated from diverse natural samples by using ethanol to kill vegetative cells and to enrich for the endospores, using nutrient medium with incubation at high temperature. Thermophiles are abundant in all samples screened. All the colonies were either *Bacillus* species or actinomycetes. About one fourth of the *Bacillus* strains isolated by incubating at 55°C could also grow at over 65°C, and the remaining three fourths could only grow at 55°C or lower temperature. The former group was identified as *B. stearothermophilus* and the latter as *B. coagulans* by using methods of conventional taxonomy. All strains isolated from the 65°C incubation temperature could grow at over 65°C and were identified as *B. stearothermophilus*.

A genomic library was constructed in *E. coli* from *B. stearothermophilus* strain XL-65-6 chromosomal DNA. This organism exhibited strong activities with cellobiose model substrates, 4-methylumbelliferyl-ß-D-glucopyranoside (MUG) and 4-methylumbelliferyl-ß-D-cellobiopyranoside (MUC).

Recombinant *E. coli* clones degrading these model substrates were readily isolated and contained genes encoding the cellobiose-specific permease of the phosphoenolpyruvate-dependent phosphotransferase system (PTS). Both MUG and MUC activities were present together and both activities were lost concurrently during subcloning experiments. A functional *E. coli ptsI* gene was required for MUC and MUG activities. The DNA fragment from *B. stearothermophilus* contained four open reading frames which appear to form a PTS *cel* operon. Intergenic stop codons for *celA*, *celB*, and *celC* overlapped the ribosomal-binding sites of the respective downstream genes. Frame-shift mutations or deletions in *celA, celB*, and *celD* were individually shown to result in a loss of MUC and MUG activities. Based upon amino acid sequence homology and hydropathy plots of translated sequences, *celA* and *celB* were identified as encoding PTS EIIB and EIIC, and *celD* encoding PTS EIIA. These translated sequences were remarkably similar to their respective *E. coli* homologues for cellobiose transport. No reported sequences exhibited a high level of homology with the *celC* gene product. The predicted carboxy-terminal region for *celC* was similar to the corresponding region of *E. coli celF*, a phospho-ß-glucosidase. An incomplete regulatory gene (*celR*) and proposed promoter sequence were located upstream of the proposed *cel* operon. A stem-loop resembling a *rho*-independent terminator was present immediately downstream from *celD*. These results indicate that

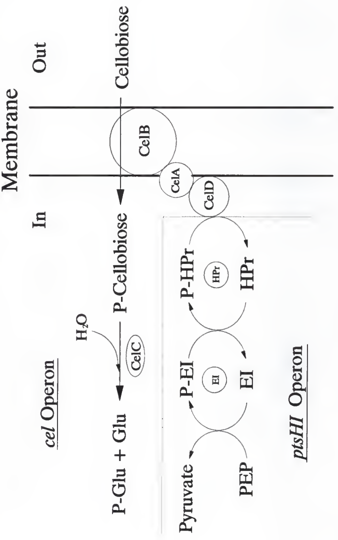
B. stearothermophilus XL-65-6 contains a cellobiose-specific PTS for cellobiose transport. Similar systems were also found in *B. subtilis* and may be common in other Gram positive bacteria. The *cel* operon may be controlled by a mechanism of transcription antitermination.

Genes encoding two general proteins of PTS, HPr (*ptsH*) and enzyme I (*ptsI*), were also cloned and sequenced from *B. stearothermophilus* XL-65-6. Both were functionally overexpressed in *E. coli*. These two genes along with the third gene, *ptsT*, form an operon, which is flanked by *ptsG* (encoding glucose permease) on the upstream and *wapA* (encoding wall associated protein) on the downstream. The *ptsH* consists of 267 nucleotides and encodes a protein of 88 amino acids with a M_r 9,166; the *ptsI* consists of 1737 nucleotides and encodes a protein of 578 amino acids with a M_r 63,470; and the *ptsT* consists of 468 nucleotides and encodes a protein of 155 amino acids with a M_r 18,428. Both HPr and enzyme I are most similar to corresponding enzymes from *B. subtilis*. The translated *ptsT* does not exhibit significant similarity to any genes in the current data base. These three genes may be translated from a common RBS which is located at upstream of *ptsH*. A polar mutation in *ptsH* almost abolished the expression of the other two genes in *E. coli*. The expression of this operon in *E. coli* at 37°C is higher than that at 30°C. *ptsT* gene product may function as an expression or activity enhancer, but its exact role is unknown. No *rho*-independent terminators were found in the regions adjacent to the *ptsHIT* operon. A transcription termination factor may be involved. Southern analysis indicates that this operon exists as a single copy in genome of *B. stearothermophilus*.

Cellobiose utilization by *E. coli* recombinants harboring the *B. stearothermophilus cel* operon was poor (slight acidification of cellobiose-MacConkey agar) although they were easily identified using chromogenic cellobiose analogues. Improved cellobiose utilization was obtained by constructing a plasmid containing both *B. stearothermophilus cel* and *ptsHI* operons. However, the growth of this recombinant in M9 minimal medium containing cellobiose as sole carbon source was not significant. The inability to grow in this medium may result from low enzymatic activities of the *B. stearothermophilus* gene products since the *ptsH* and *ptsI* products were abundant.

A general model for initiating cellobiose utilization in *B. stearothermophilus* has been proposed (Fig. 15). This organism uses PTS for cellobiose uptake from the environment and to initiate its metabolism. Genes encoding the cellobiose-specific permease form an operon, consisting of four genes, *celA, celB, celC,* and *celD*. CelB is the major membrane spanning protein, CelA and D are accessory proteins. These 3 proteins along with enzyme I and HPr form the transport system. Cellobiose is transported into cytoplasm by this system and concurrently phosphorylated. The

Figure 15. Model for cellobiose transport and metabolism in *B. stearothermophilus*. EI and HPr are encoded by *ptsI* and *ptsH*. The cellobiose-specific permease is encoded by *celA*, *celB*, and *celD*. The *celC* gene product is proposed to encode a cleavage enzyme for cellobiose-phosphate.


Membrane
Out
In
Cellobiose
CelB
CelA
CelD
P-Cellobiose
H_2O
CelC
P-Glu + Glu
cel Operon
P-HPr
HPr
HPr
P-EI
EI
EI
Pyruvate
PEP
ptsHI Operon

phosphorylated cellobiose is then hydrolyzed by CelC with the product entering the glycolytic pathway.

LITERATURE CITED

Al-Awadhi, N., T. Egli, G. Hamer, and E. Wehrli. 1989. Thermotolerant and thermophilic solvent-utilising methylotrophic, aerobic bacteria. System. Appl. Microbiol. 11:207-216.

Alfred, C.-A. and B. Chassy. 1988. Molecular cloning and nucleotide sequence of the factor III^{lac} gene of *Lactobacillus casei*. Gene 62:277-288.

Alfred, C.-A. and B. Chassy. 1990. Molecular cloning and DNA sequence of *lac*E, the gene encoding the lactose-specific enzyme II of the phosphotransferase system of *Lactobacillus casei*. J. Biol. Chem. 265:22561-22568.

Alpert, C. A., M. Dörschug, D. Saffen, R. Frank, J. Deutscher, and W. Hengstenberg. 1985. The bacterial phosphoenolpyruvate-dependent phosphotransferase system. Isolation of active site peptides by reversed-phase high performance liquid chromatography and determination of their primary structure. J. Chromatogr. 326:363-371.

Al-Zaag, A. 1989. Molecular cloning of cellobiose and other ß-glucosidase determinants from *Klebsiella oxytoca*. J. Biotechnol. 12:79-86.

Amelunxen, R. and M. Lins. 1968. Comparative thermostability of enzymes from *Bacillus stearothermophilus* and *Bacillus cereus*. Arch. Biochem. Biophys. 125:765-769.

Amelunxen, R. E. and A. L. Murdock. 1978. Mechanisms of thermophily, p. 343-393. *In* A. I. Laskin and H. Lechevalier (ed.), Critical Reviews in Microbiology, vol 6. CRC Press, Inc.

Aragno, M. 1978. Enrichment, isolation and preliminary characterization of a thermophilic. endospore-forming hydrogen bacterium. FEMS Microbiol. Lett. 3:13-15.

Aragno, M. 1992. Aerobic, chemolithoautotrophic, thermophilic bacteria, p. 77-103. *In* J. K. Kristjansson (ed.), Thermophilic Bacteria. CRC Press, Boca Raton, Florida.

Arnaud, M., P. Vary, M. Zagorec, A. Klier, M. Debarbouille, P. postma, and G. Rapoport. 1992. Regulation of the *sacPA* operon of *Bacillus subtilis*: identification of phosphotransferase system components involved in SacT activity. J. Bacteriol. 174:3161-3170.

Atlas, R. M. and L. C. Parks (ed.) 1993. Handbook of Microbiological Media. CRC Press, Inc.

Bartolomeo, A. D., F. Trotta, F. L. Rosa, G. Saltalamacchia, and V. Mastrandrea. 1991. Numerical analysis and DNA base compositions of some thermophilic *Bacillus* species. Int. J. Syst. Bacteriol. 41:502-509.

Benson, D., D. J. Lipman, and J. Ostell. 1993. GenBank. Nucleic Acids Res. 21:2963-2965.

Böhm, G. and R. Jaenicke. 1994. Relevance of sequence statistics for the properties of extremophilic proteins. Int. J. Peptide Protein Res. 43:97-106.

Boyd, D. A., D. G. Cvitkovitch, and I. R. Hamilton. 1994. Sequenc and expression of the genes for HPr (*ptsH*) and Enzyme I (*ptsI*) of the Phosphoenolpyruvate-dependent phosphotransferase transport system from *Streptococcus mutans*. Infect. Immun. 62:1156-1165.

Breidt, F., Jr., W. Hengstenberg, U. Finkeldei, and G. C. Stewart. 1987. Identification of the gene for the lactose-specific components of the phosphotransferase system in the *lac* operon of *Staphylococcus aureus*. J. Biol. Chem. 262:16444-16449.

Brock, T. D. 1986. Introduction: An Overview of The Thermophiles, p. 1-16. *In* T. D. Brock (ed.), Thermophiles: General, Molecular, and Applied Microbiology. John Wiley & Sons, Inc.

Brock, T. D. and M. T. Madigan (ed.). 1991. Biology of Microorganisms, 6th ed. Prentice Hall, Englewood Cliffs, New Jersey.

Brown, B. J., J. F. Preston, and L. O. Ingram. 1991. Cloning of alginate lyase gene (*alkM*) and expression in *Escherichia coli*. Appl. Environ. Microl. 57:1870-1872.

Bruchhaus, I and E. Tannich. 1993. Primary structure of the pyruvate phosphate dikinase in *Entamoeba histolytica*. Mol. Biochem. Parasitol. 62:153-156.

Buswell, J. A. 1975. Metabolism of phenol and cresols by *Bacillus stearothermophilus*. J. Bacteriol. 124:1077-1083.

Buswell, J. A. and D. G. Twomey. 1975. Utilization of phenol and crestols by *Bacillus stearothermophilus*, strain PH24. J. Gen. Microbiol. 87:377-379.

Byrne, C. R., R. S. Monroe, K. A. Ward, and N. M. Kredich. 1988. DNA sequences of the *cysK* regions of *Salmonella typhimurium* and *Escherichia coli* and linkage of the *cysK* regions to *ptsH*. J. Bacteriol. 170:3150-3157.

Cato, E. P., W. L. George and S. M. Finegold. 1986. Genus *Clostridium* Prazmowski 19880, 23, p. 1141-1200. *In* P. H. A. Sneath, N. S. Mair, M. E. Sharpe and J. G. Hold (ed.), Bergey's Manual of Systematic Bacteriology, Vol 2. Williams & Wilkins, Baltimore, MD.

Chen, Y., W. J. Fairbrother, and P. E. Wright. 1993. Three-dimensional structures of the central regulatory proteins of the bacterial phosphotransferase system, HPr and enzyme IIA^{glc}. J. Cell. Biochem. 51:75-82.

Claus D. and R. C. W. Berkeley. 1986. Genus *Bacillus* Cohn 1872, 174, p. 1105-1139. *In* P. H. A. Sneath, N. S. Mair, M. E. Sharpe and J. G. Hold (ed.), Bergey's Manual of Systematic Bacteriology, Vol 2. Williams & Wilkins, Baltimore, MD.

Coughlan, M. P. 1985. Cellulose hydrolysis: The potential, the problems and relevant research at Galway. Biochem. Soc. Trans. 13:405-406.

Coughlan, M. P. 1988. Staining techniques for the detection of the individual components of cellulolytic enzyme system. *In* T. M. Wood and S. T. Kellogg (ed.), Methods Enzymol. 160:135-144.

Coughlan, M. P. and F. Mayer. 1992. The Cellulose-Decomposing Bacteria and Their Enzyme Systems. P. 460-516. *In* A. Balows, H. G. Trüper, M. Dworkin, W. Harder, and K. Schleifer (ed.), The Prokaryotes, 2nd ed. A Handbook on the Biology of Bacteria: Ecophysiology, Isolation, Identification, Applications, Vol. I. Springer-Verlag.

Cutting, S. M. and P. B. Vander Horn. 1990. Genetic Analysis, p. 27-74. *In* C. R. Harwood and S. M. Cutting (ed.), Molecular biological methods for *Bacillus*. John Wiley & Sons Ltd., Chichester, England.

Darland, G and T. D. Brock. 1971. *Bacillus acidocaldarius* sp. nov., an acidophilic thermophilic spore-forming bacterium. J. Gen. Microbiol. 67:9-15.

Davies, G. J. J. A. Littlechild, H. C. Watson and L. Hall.

1991. Sequence and expression of the gene encoding 3-phosphoglycerate kinase from *Bacillus stearothermophilus*. Gene 109:39-45.

Debarbouille, M., M. Arnaud, A. Fouet, A. Klier and G. Rapoport. 1990. The *sacT* gene regulating the *sacPA* operon in *Bacillus subtilis* shares strong homology with transcriptional antiterminators. J. Bacteriol. 172:3966-3973.

Débarbouillé, M., I. Martin-Verstraete, A. Klier and G. Rapoport. 1991. The transcriptional regulator levR of *Bacillus subtilis* has domains homologous to both σ^{54}- and phosphotransferase system-dependent regulators. Proc. Natl. Acad. Sci. U.S.A. 88:2212-2216.

De Reuse, H. and A. Danchin. 1988. The *ptsH, ptsI*, and *crr* genes of the *Escherichia coli* phosphoenolpyruvate-dependent phosphotransferase system: a complex operon with several modes of transcription. J. Bacteriol. 170:3827-3837.

De Reuse, H. and A. Danchin. 1991. Positive regulation of the *pts* operon of *Escherichia coli*: genetic evidence for a signal transduction mechanism. J. Bacteriol. 173:727-733.

De Reuse, H., A. Kolb, and A. Danchin. 1992. Positive regulation of the expression of the *Escherichia coli pts* operon. Identification of the regulatory regions. J. Mol. Biol. 226:623-635.

Deutscher, J., J. Reizer, C. Fischer, A. Galinier, M. H. Saier, Jr., and M. steinmetz. 1994. Loss of protein kinase-catalyzed phosphorylation of HPr, a phosphocarrier protein of the phosphotransferase system, by mutation of the *ptsH* gene confers catabolite repression resistance to several catabolic genes of *Bacillus subtlis*. J. Bacteriol. 176:3336-3344.

Devereux, J., P. Haeberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12:387-395.

de Vos, W. M., I. Boerrigter, R. J. van Rooyen, B. Reiche and W. Hengstenberg. 1990. Characterization of the lactose-specific enzymes of the phosphotransferase system in *Lactococcus lactis*. J. Biol. Chem. 265:22554-22560.

Dijkhuizen, L., N. Arfman, M. M. Attwood, A. G. Brooke, W. Harder, and E. M. Watling. 1988. Isolation and initial characterization of thermotolerant methylotrophic *Bacillus* strains. FEMS Microbiol. Lett. 52:209-214.

DiRuggiero, J., F. T. Robb, R. Jagus, H. H. Klump, K. M. Borges, M. Kessel, X. Mai, and M. W. W. Adams. 1993. Characterization, cloning, and *in vitro* expression of the extremely thermostable glutamate dehydrogenase from the hyperthermophilic Archaeon, ES4. J. Biol. Chem. 268:17767-17774.

Drlica, K. and J. Rouviere-Yaniv. 1987. Histone-like proteins of bacteria. Microbiol. Rev. 51:301-319.

Droffner, M. L. and N. Yamamoto. 1985. Isolation of thermophilic mutants of *Bacillus subtilis* and *Bacillus pumilus* and transformation of the thermophilic trait to mesophilic strains. J. Gen. Microbiol. 131:2789-2794.

El Hassouni, M., B. Henrissat, M. Chippaux and F. Barras. 1992. Nucleotide sequences of the *arb* genes, which control ß-glucoside utilization in *Erwinia chrysanthemi*: comparison with the *Escherichia coli bgl* operon and evidence for a new ß-glycohydrolase family including enzymes from eubacteria, archeabacteria, and humans. J. Bacteriol. 174:765-777.

Ferrari, E., D. J. Henner, and M. Y. Yang. 1985. Isolation of an alanine racemase gene from *Bacillus subtilis* and its use for plasmid maintenance in *B. subtilis*. Biotechnology (N.Y.) 3:1003-1007.

Foster, S. J. 1993. Molecular analysis of three major wall-associated proteins of *Bacillus subtilis* 168: evidence for processing of the product of a gene encoding a 258 kDa precursor, two-domain ligand-binding protein. Mol. Microbiol. 8:299-310.

Fouet, A., M. Arnaud, A. Klier and G. Rapoport. 1987. *Bacillus subtilis* sucrose-specific enzyme II of the phosphotransferase system: expression in *Escherichia coli* and homology to enzyme II from enteric bacteria. Proc. Natl. Acad. Sci. 84:8773-8777.

Fouet, A., M. Arnaud, A. Klier, and G. Rapoport. 1989. Genetics of the phosphotransferase system of *Bacillus subtilis*. FEMS Microbiol. Rev. 63:175-182.

Fraenkel, D. G., F. Falcoz-Kelly, and B. L. Horecker. 1964. The utilization of glucose 6-phosphate by glucokinaseless and wild-type strains of *Escherichia coli*. Proc. Natl. Acad. Sci. 52:1027-1213.

Frazier, C. F. 1958. Food Microbiology. McGraw-Hill, New York.

Gagnon, G., C. Vadeboncoeur, R. C. Levesque, and M.

Frenette. 1992. Cloning, sequencing, and expression in *Escherichia coli* of the *ptsI* gene encoding enzyme I of the phosphoenolpyruvate:sugar phosphotransferase transport system from *Streptococcus salivarius*. Gene 121:71-78.

Galakatos, N. G., E. Daub, D. Botstein, and C. T. Walsh. 1986. Biosynthetic *alr* alanine racemase from *Salmonella typhimurium*: DNA and protein sequence determination. Biochem. 25:3255-3260.

Gonzy-Tréboul, G., M. Zagorec, M.-C. Rain-Guion and M. Steinmetz. 1989. Phosphoenolpyruvate:sugar phosphotransferase system of *Bacillus subtilis*: nucleotide sequence of *ptsX, ptsH* and the 5'-end of *ptsI* and evidence for a *ptsHI* operon. Mol. Microbiol. 3:103-112.

Hall, B. G. and W. Faunce III. 1987. Functional genes for cellobiose utilization in natural isolates of *Escherichia coli*. J. Bacteriol. 169:2713-2717.

Hall, B. G. and L. Xu. 1992. Nucleotide sequence, function, activation, and evolution of the cryptic *asc* operon of *Escherichia coli* K12. Molecular Biol. Evol. 9:688-706.

Harwood, C. R. 1989. Introduction to the biotechnology of *Bacillus*. p. 1-4. *In* C. R. Harwood (ed.), *Bacillus*. Plenum Press.

Hengstenberg, W., B. Reiche, R. Eisermann, R. Fischer, U. Kebler, A. Tarrach, W. M. De Vos, H.-R. Kalbitzer and S. Glaser. 1989. Structure and function of proteins involved in sugar transport by the PTS of Gram-positive bacteria. FEMS Microbiol. Rev. 63:35-42.

Higgins, D. G., A. J. Bleasby and R. Fuchs. 1992. CLUSTAL V: improved software for multiple sequence alignment. Comput. Appl. Biosci. 8:189-191.

Ishikawa, K., S. Kimura, S. Kanaya, K. Morikawa, and H. Nakamura. 1993. Structural study of mutants of *Escherichia coli* ribonuclease HI with enhanced thermostability. Protein Engineering 6:85-91.

Klein, P., M. Kanehisa and C. Delisi. 1985. The detection and classification of membrane spanning proteins. Biochim. Biophys. Acta 815:468-476.

Kluepfel, D. 1988. Screening of prokaryotes for cellulose- and hemicellulose-degrading enzymes. *In* T. M. Wood and S. T. Kellogg (ed.), Methods Enzymol. 160:180-186.

Koffler, H. and G. O. Gale. 1957. The relative thermostability of cytoplasmic proteins from thermophilic bacteria. Arch. Biochem. Biophys. 67:249-251.

Kohlbrecher, D., R. Eisermann, and W. Hengstenberg. 1992. Staphylococcal phosphoenolpyruvate-dependent phosphotransferase system: molecular cloning and nucleotide sequence of the *Staphylococcus carnosus ptsI* gene and expression and complementation studies of the gene product. J. Bacteriol. 174:2208-2214.

Koransky, J. R., S. D. Allen, and V. R. Dowell, Jr. 1978. Use of ethanol for selective isolation of sporeforming microorganisms. Appl. Environ. Microbiol. 35:762-765.

Kristjansson, J. K. and K. O. Stetter. 1992. Thermophilic Bacteria, p. 1-18. *In* J. K. Kristjansson (ed.), Thermophilic Bacteria. CRC Press, Boca Raton, Florida.

Krohn, B. M. and J. A. Lindsay. 1991. Purification and characterization of a thermostable α-glucosidase from a *Bacillus subtilis* high-temperature growth transformant. Curr. Microbiol. 22:273-278.

Kukuruzinska, M. A., W. F. Harrington, and S. Roseman. 1982. Sugar transport by the bacterial phosphotransferase system. Studies on the molecular weight and association of Enzyme I. J. Biol. Chem. 257:14470-14476.

Kundig, W., S. Ghosh, and S. Roseman. 1964. Phosphate bound to histidine in a protein as an intermediate in a novel phosphotransferase system. Proc. Natl. Acad. Sci. U.S.A. 52:1067-1074.

Kyte, J and R. F. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157:105-132.

Lengeler, J. W. and H. Steinberger. 1978. Analysis of the regulatory mechanisms controlling the activity of the hexitol transport systems in *Escherichia coli* K12. Mol. Gen. Genet. 164:163-169.

Lèvy, S., H. De Reuse, and A. Danchin. 1989. Antisense expression at the *ptsH-ptsI* locus of *Escherichia coli*. FEMS Microbiol. Lett. 57:35-38.

LiCalsi, C., T. S. Crocenzi, E. Freire, and S. Roseman. 1991. Sugar transport by the bacterial phosphotransferase system. Structural and thermodynamic domains of enzyme I of *Salmonella typhomurium*. J. Biol. Chem. 266:19519-19527.

Lindsay, J. A. and E. H. Creaser. 1975. Enzyme thermostability is a transformable property between *Bacillus* spp. Nature 255:650-652.

Lobocka, M.B., J. Hennig, J. Wild, and T. Klopotowski. 1992. Structure and expression of the *dad* operon of *Escherichia coli* K12, coding for the D-amino acid dehydrogenase and the catabolic alanine racemase. GenBank.

Mattoo, R. L. and E. B. Waygood. 1983. Determination of the levels of HPr and enzyme I of the phosphoenolpyruvate-sugar phosphotransferase system in *Escherichia coli* and *Salmonella typhimurium*. Can. J. Biochem. Cell Biol. 61:29-37.

McLinden, J., A. Murdock, and R. Amelunxen. 1986. Mechanism of thermophily for thermolabile glyceraldehyde-3-phosphate dehydrogenase from the facultative thermophile *Bacillus coagulans* KU. Biochim. Biophys. Acta 871:207-216.

Meadow, N. D., D. K. Fox, and S. Roseman. 1990. The bacterial phosphoenolpyruvate: glycose phosphotransferase system. Annu. Rev. Biochem. 59:497-542.

Merkler, D. J., K. Srikumar, S. P. Marchese-Ragona, and F. C. Wedler. 1988. Aggregation and thermo-inactivation of glutamine synthetase from an extreme thermophile, *Bacillus caldolyticus*. Biochim. Biophys. Acta 952:101-114.

Minnikin, D. E. and M. Goodfellow. 1981. Lipids in the classification of *Bacillus* and related taxa, p. 59-90. *In* R. C. W. Berkeley and M. Goodfellow (ed.), The Aerobic Endospore-Forming Bacteria: Classification and Identification. Academic Press, Inc., London.

Misset, O., M. Brouwer, and G. T. Robillard. 1980. *Escherichia coli* phosphoenolpyruvate-dependent phosphotransferase system. Evidence that the dimmer is the active form of Enzyme I. Biochem. 19:883-890.

Mitchell, W. J., D. W. Saffen, and S. Roseman. 1987. Sugar transport by the bacterial phosphotransferase system: *in vivo* regulation of lactose transport in *Escherichia coli* by III^{Glu}, a protein of the phosphoenolpyruvate: glucose phosphotransferase system. J. Biol. Chem. 262:16254-16260.

Moran, C. P., Jr. 1993. RNA polymerase and transcription factors. *In*: A. L. Sonenshein, J. A. Hoch and R. Losick (ed.) *Bacillus subtilis* and Other Gram-Positive Bacteria: Biochemistry, Physiology, and Molecular Genetics. American Society for Microbiology.

Mozhaev, V. V. and K. Martinek. 1984. Structure-stability relationships in proteins: new approaches to stabilizing enzymes. Enz. Microb. Technol. 6:50-59.

Murphy, N., D. J. McConnell, and B. A. Cantwell. 1984. The DNA sequence of the gene and genetic control sites for the excreted *Bacillus subtilis* enzyme ß-glucanase. Nucl. Acids Res. 12:5355-5367.

Neidhardt, F. C., J. L. Ingraham, and M. Schaechter. 1990. Physiology of the Bacterial Cell: A Molecular Approach. Sinauer Associates, Inc. Publishers, Sunderland, Massachusetts.

Niersbach, M., F. Kreuzaler, R. H. Geerse, P. W. Postma, and H. J. Hirsch. 1992. Cloning and nucleotide sequence of the *Escherichia coli* K-12 *psa* gene, encoding PEP synthase. Mol. Gen. Genet. 231:332-336.

Ogasawara, N. and H. Yoshikawa. 1992. Genes and their organization in replication origin region of bacterial chromosome. Mol. Microbiol. 6:629-634.

Parker, L. L. and B. G. Hall. 1990. Characterization and nucleotide sequence of the cryptic *cel* operon of *Escherichia coli* K12. Genetics 124:455-471.

Pocalyko, D. J., L. J. Carroll, B. M. Martin, P. C. Babbitt, and D. Dunaway-Mariano. 1990. Analysis of sequence homologies in plant and bacterial pyruvate phosphate dikinase, Enzyme I of the bacterial phosphoenolpyruvate:sugar phosphotransferase system and other PEP-utilizing enzymes. Identification of potential catalytic and regulatory motifs. Biochemistry 29:10757-10765.

Postma, P. W., J. W. Lengeler, and G. R. Jacobson. 1993. Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. Microbiol. Rev. 57:543-594.

Pries, A., H. Priefert, N. Krüger and A. Steinbüchel. 1991. Identification and characterization of two *Alcaligenes eutrophus* gene loci relevant to the poly(ß-hydroxybutyric acid)-leaky phenotype which exhibit homology to *ptsH* and *ptsI* of *Escherichia coli*. J. Bacteriol. 173:5843-5853.

Priest, F. G. 1989. Isolation and identification of aerobic endospore-forming bacteria, p. 27-56. *In* C. R. Harwood (ed.), *Bacillus*. Plenum Press, New York.

Reizer, J., C. Hoischen, A. Reizer, and M. H. Saier. 1993. Sequence analysis and evolutionary relationships among the energy coupling proteins, Enzyme I and Hpr of bacterial phosphoenolpyruvate: sugar phosphotransferase system. Protein Sci. 2:506-521.

Reizer, J., A. Reizer and M. H. Saier, Jr. 1990. The cellobiose permease of *Escherichia coli* consists of three proteins and is homologous to the lactose permease of *Staphylococcus aureus*. Res. Microbiol. 141:1061-1067.

Reizer, J., A. H. Romano, and J. Deutscher. 1993. The role of phosphorylation of HPr, a phosphocarrier protein of the phosphotransferase system, in the regulation of carbon metabolism in Gram-positive bacteria. J. Cell. Biochem. 51:19-24.

Rentier-Delrue, F. S. C. Mande, S. Moyens, P. Terpstra, V. Mainfroid, K. Goraj, M. Lion, W. G. J. Hol, and J. A. Martial. 1993. Cloning and overexpression of the triosephosphate isomerase genes from psychrophilic and thermophilic bacteria: structural comparison of the predicted protein sequences. J. Mol. Biol. 229:85-93.

Richterich, P., N. Lakey, G. Gryan, L. Jaehn, L. Mintz, K. Robison, and G. M. Church. 1994. Automated multiplex sequencing of the *Escherichia coli* genome. GenBank (Accession Number Uoooo7).

Roberts, R. J. and D. Macelis. 1992. Restriction enzymes and their isoschizomers. Nucleic Acids Res. 20:2167-2180 (Suppl.).

Rosenberg, M., and D. Court. 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Ann. Rev. Genet. 13:319-353.

Roy, A., C. F. Lu, D. L. Marykwas, P. Lipke, and J. Kurjan. 1991. The AGA1 product is involved in cell surface attachment of the *Saccharomyces cervisiae* cell adhesion glycoprotein a-agglutinin. GenBank.

Ryu, S. and S. Garges. 1994. Promoter switch in the *Escherichia coli pts* operon. J. Biol. Chem. 269:4767-4772.

Saffen, D. W., K. A. Presper, T. L. Doering, and S. Roseman. 1987. Sugar transport by the bacterial phosphotransferase system-molecular cloning and structural analysis of the *Escherichia coli ptsH, ptsI*, and crr genes. J. Biol. Chem. 262:16241-16253.

Saier, M. H., Jr., and J. Reizer. 1992. Proposed uniform nomenclature for the proteins and protein domains of the bacterial phosphoenolpyruvate:sugar phosphotransferase system. J. Bacteriol. 174:1433-1438.

Saier, M. H., Jr. and M. Yamada. 1987. Evolutionary considerations concerning the bacterial phosphoenolpyruvate: sugar phosphotransferase system. *In* J. Reizer and A. Peterkofsky (ed.), Sugar Transport and Metabolism in Gram-Positive Bacteria. Ellis Horwood Limited, Chichester, England.

Saier, M. H., Jr., M. Yamada, J. Lengeler, B. Erni, K. Suda, P. Argos, K. Schnetz, B. Rak, C. A. Lee, G. C. Stewart, K. G. Peri, and E. B. Waygood. 1988. Sugar permeases of the bacterial phosphoenolpyruvate-dependent phosphotransferase system: sequence comparisons. FASEB J. 2:199-208.

Sakoda, H. and T. Imanaka. 1992. Cloning and sequencing of the gene coding for alcohol dehydrogenase of *Bacillus stearothermophilus* and rational shift of the optimum pH. J. Bacteriol. 174:1397-1402.

Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.

Sharp. R. J., P. W. Riley, and D. White. 1992. Heterotrophic Thermophilic *Bacillus*, P. 19-50. *In* J. K. Kristjansson (ed.), Thermophilic Bacteria. CRC Press, Boca Raton.

Schägger, H. and G. van Jagow. 1987. Tricine-sodium dodecyl sulfate-polyacylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Analytical Biochem. 166:368-379.

Schenk, A and M. Aragno. 1979. *Bacillus schlegelii*, a new species of thermophilic, facultatively chemolithoautotrophic bacterium oxidizing molecular hydrogen. J. Gen. Microbiol. 115:333-341.

Sei, T., C. T. Chang, H. Mikami, and V. Oshima. 1978. Deoxyribonucleic acid homology and taxonomy of the genus *Bacillus*. Int. J. Syst. Bacteriol. 28:182-

Sidler, W. and H. Zuber. 1977. The production of extracellular thermostable neutral proteinase and α-amylase by *Bacillus stearothermophilus*. Eur. J. Appl. Microbiol. 4:255.

Skepecky, R. A. and H. E. Hemphill. 1992. The Genus *Bacillus* -Nonmedical. p. 1663-1696. *In* A. Balows, H. G. Trüper, M. Dworkin, W. Harder, and K. Schleifer (ed.), The Prokaryotes, 2nd ed. A handbook on the biology of bacteria: ecophysiology, isolation, identification, applications, Vol. II. Springer-Verlag.

Smith, N. R., R. E. Gordon, and F. E. Clark. 1952. Aerobic spore-forming bacteria. U. S. Department of Agriculture, Washington, D.C., Monogr. No. 16.

Ståhl, S. R. 1991. Plasmids in *Bacillus stearothermophilus* coding for bacteriocinogeny and temperature resistance. Plasmid 26:94-107.

Stewart, G. C. 1993. Catabolite repression in the Gram-positive bacteria: generation of negative regulators of transcription. J. Cell. Biochem. 51:25-28.

Stock, J. B., E. B. Waygood, N. D. Meadow, P. W. Postma, and S. Roseman. 1982. Sugar transport by the bacterial phosphotransferase system. The glucose receptors of the *Salmonella typhimurium* phosphotransferase system. J. Biol. Chem. 257:14543-14552.

Sundaram, T. K. 1986. Physiology and growth of thermophilic bacteria, p. 75-106. *In* T. D. Brock (ed.), Thermophiles: General, Molecular, and Applied Microbiology. John Wiley & Sons, Inc.

Tajima, M., L. Urabe, K. Yutani, and H. Okada. 1976. Role of calcium ions in the thermostability of thermolysin and *Bacillus subtilis* var. *amylosacchariticus* neutral protease. Eur. J. Biochem. 64:243-247.

Tanizawa, K., A. Ohshima, A. Scheidegger, K. Inagaki, H. Tanaka, and K. Soda. 1988. Thermostable alanine racemase from *Bacillus stearothermophilus*: DNA and protein sequence determination and secondary structure prediction. Biochem. 27:1311-1316.

Titgemeyer, F., R. Eisermann, W. Hengstenberg, and J. W. Lengeler. 1990. The nucleotide sequence of *ptsH* gene from *Klebsiella pneumoniae*. Nucleic Acids Res. 18:1898.

Vadeboncoeur, C., L. Gauthier, G. Gagnon, A. Leduc, D. Brochu, R. Lapointe, B. Desjardins, and M. Frenette. 1994. Properties of a *Streptococcus salvarius* spontaneous mutant in which the methionine at position 48 in the protein HPr has been replaced by a valine. J. Bacteriol. 176:524-527.

Van Charldorp, R., A. M. Van Kimmenade, and P. H. Van

Knippenberg. 1981. Sequence and secondary structure of the colicin fragment of *Bacillus stearothermophilus* 16S ribosomal RNA. Nucleic Acids Res. 9:4909-4917.

van der Vlag, J., K. van Dam, and P. W. Postma. 1994. Quantification of the regulation of glycerol and maltose metabolism by IIA^{Glc} of the phosphoenolpyruvate-dependent glucose phosphotransferase system in *Salmonella typhimurium*. J. Bacteriol. 176:3518-3526.

Weigel, N., M. A. Kukuruzinska, A. Nakazawa, E. B. Waygood, and S. Roseman. 1982. Sugar transport by bacterial phosphotransferase system. Phosphoryl transfer reactions catalyzed by Enzyme I of *Salmonella typhimurium*. J. Biol. Chem. 257:14477-14491.

Welker, N. E. 1976. Microbial endurance and resistance to heat stress. 20th Symp. Soc. Gen. Microbiol. Cambridge University Press.

Welker, N. E. 1993. The genetic map of *Bacillus stearothermophilus* NUB36, p. 483-487. *In* A.L. Sonenshein, J. A. Hoch, and R. Losick (ed.), *Bacillus subtilis* and Other Gram-Positive Bacteria-Biochemistry, Physiology, and Molecular Genetics. American Society for Microbiology, Washington, DC.

Williams, R. A. D. 1992. The genus *Thermus*, p. 51-62. *In* J. K. Kristjansson (ed.), Thermophilic Bacteria. CRC Press, Inc. Boca Raton, Florida.

Wolf, J. and A. N. Barker. 1968. The genus *Bacillus*: aids to the identification of its species, p. 93-109. *In* M. Gibbs and F. A. Skinner (ed.), Identification Methods for Microbiologists, part B. Academic Press, Inc., London.

Wolf, J. and R. J. Sharp. 1981. Taxonomic and related aspects of thermophiles within the genus *Bacillus*, p. 251-296. *In* R. C. W. Berkeley and M. Goodfellow (ed.), The Aerobic Endospore-Forming Bacteria: Classification and Identification. Academic Press, Inc., London.

Wood, T. M. and K. M. Bhat. 1988. Methods for measuring cellulase activities. *In* W. A. Wood and S. T. Kellogg (ed.), Methods Enzymol. 160:87-112.

Wu, L.-F., J. M. Tomich, and M. H. Saier, Jr. 1990. Structure and evolution of a multidomain multiphosphoryl transfer protein. Nucleotide sequence of the *fruB(HI)* gene in *Rhodobacter capsulatus* and comparisons with homologous genes from other organisms. J. Mol. Biol. 213:687-703.

Ye, J. J., J. W. Neal, X. Cui, J. Reizer, and M. H. Saier, Jr. 1994a. Regulation of the glucose:H^+ symporter by metabolite-activated ATP-dependent phosphorylation of HPr in *Lactobacillus brevis*. J. Bacteriol. 176:3484-3492.

Ye, J. J., J. Reizer, X. Cui, and M. H. Saier, Jr. 1994b. Inhibition of the phosphoenolpyruvate:lactose phosphotransferase system and activation of a cytoplasmic sugar-phosphate phosphatase in *Lactococcus lactis* by ATP-dependent metabolite-activated phosphorylation of serine 46 in the phosphocarrier protein HPr. J. Biol. Chem. 269:11837-11844.

Yutani, K. 1976. Role of calcium ions in the thermostability of amylase produced by *Bacillus stearothermophilus*, p. 91-103. *In* H. Zuber (ed.), Enzymes and Proteins from Thermophilic Microorganisms-Structure and Function. Birkäuser Verlag, Basel.

Zagorec, M. and P. W. Postma. 1992. Cloning and nucleotide sequence of the *ptsG* gene of *Bacillus subtilis*. Mol. Gen. Genet. 234:325-328.

Zamost, B. L., H. K. Nielsen and R. L. Starnes. 1991. Thermostable enzymes for industrial applications. J. Indust. Microbiol. 8:71-82.

Zhu, P, J. Reizer, A. Reizer, and A. Peterkofsky. 1993. Unique monocistronic operon (*ptsH*) in *Mycoplasma capricolum* encoding the phosphocarrier protein, HPr, of the phosphoenolpyruvate:sugar phosphotransferase system. J. Biol. Chem. 268:26531-26540.

Zukowski, M. M., L. Miller, P. Cosgwell, K. Chen, S. Aymerich, and M. Steinmetz. 1990. Nucleotide sequence of the *sacS* locus of *Bacillus subtilis* reveals the presence of two two regulatory genes. Gene 90:153-155.

BIOGRAPHICAL SKETCH

Xiaokuang Lai was born on September 27, 1960, in Hangzhou, China. After graduating from The Second High School of Hangzhou in 1978, he attended Zhejiang Agricultural University and received his Bachelor of Science degree in horticulture in 1982. After he worked as a research associate in Zhejiang Academy of Agricultural Sciences for five years, he was awarded a research assistantship from the University of Hawaii at Manoa in 1987 and went to the beautiful Aloha state and received his Master of Science degree in horticulture with Dr. Richard W. Hartmann in 1990. In the same year, he began his doctoral studies at the University of Florida. After one semester enrolled in the Department of Horticulture, he transferred to the Department of Microbiology and Cell Science and worked in Dr. Lonnie O. Ingram's lab. He learned amazing microbiology and molecular biology there.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Lonnie O'Neal Ingram

Lonnie O'Neal Ingram
Chairman, Professor of
Microbiology and Cell Science

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

John E. Gander

John E. Gander
Professor of Microbiology and
Cell Science

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

K.T. Shanmug

Keelnatham T. Shanmugam
Professor of Microbiology and
Cell Science

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Francis C Davis

Francis C. Davis
Associate Professor of
Microbiology and Cell Science

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

James A Lindsay

James A. Lindsay
Professor of Food Science
and Human Nutrition

This dissertation was submitted to the Graduate Faculty of the College of Agriculture and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

August 1994

Jack L. Fry

Dean, College of
Agriculture

Dean, Graduate School

LD
1780
1994
.L185

UNIVERSITY OF FLORIDA
3 1262 08554 8872